TECHNISCHE UNIVERSITÄT DRESDEN

Faculty of Environmental Sciences Institute of Cartography

MASTER THESIS

Historical Spatio-temporal data in current GIS
Case Study: German-Herero war of resistance 1904

submitted by

Susanna Ambondo Abraham

13 July 1991, Oshikuku

Matriculation Number: 4576256

International Master of Science in Cartography

Technical University of Dresden

Supervisors:
First Supervisor: Prof. Lars Bernard
Second Supervisor: Dr.-Ing. Christian Murphy

Submission date: 09 October 2017

Statement of Authorship

I hereby declare that I am the sole author of this master thesis and information directly or indirectly taken from other sources have been noted as such. I further declare that I have not submitted this thesis nor any partial work of this thesis to an examination committee.

Dresden, 09 October 2017

Acknowledgements

First of all, I would like to thank the Lord for his provision, strength and comfort on this journey. I could have not done it without his grace and mercy upon me. This has been the most challenging and emotional episode of my life. I would like to take this time and express my appreciation to all that helped me make my studies in Germany a worthwhile. I would like to express my sincere gratitude to Prof. Dr. Lars Bernard for agreeing to supervise me and for his guidance on this thesis. I would also like to thank Dr-Ing Christian Murphy my second supervisor and Prof Dirk Burghardt for serving on my thesis committee and for their kind help and time. A special appreciation goes to Mrs Juliane Cron, for her kind assistance from the very first day on this master's course, for motivating me to not give up when I felt like leaving. I am really grateful for all that you have done for me.

Another special thanks goes to Mr Ankama Nadagho for his continuous support and for always believing in me. Equally, I would like to thank my dear friend Donald Akogou for being there to review my work and for always listening to my issues. My dear friend Maja Kalinic, you have not only been a classmate but also a sister to me. I am grateful to have met you and for making my life in Germany less stressful.

Finally, my dear mama and sister for being my motivation to take on this amazing challenge. I thank you for your support and love.

Abstract

The growing developments in Geographic Information Retrieval (GIR) and Temporal Information Retrieval (TIR) techniques have given new ways to explore digital text archives for spatio-temporal data. Often, we are faced with questions regarding past events and the answers are hidden in the historical text archives. The question is how to retrieve the answers from the text documents. This thesis aims to contribute to the better understanding of spatio-temporal information extraction from text documents. Natural Language Processing techniques were used to develop an information extraction framework using GATE development framework. The developed framework uses gazetteer matching and pattern based rules to recognize and annotate elements in the documents. The goal is to use the extracted spatio-temporal data to study the time-geography context of the German-Herero war of resistance by using GIS. However, there are shortcomings regarding spatio-temporal functionalities in GIS applications. Space-time cube analysis, trajectory analysis and spatio-temporal clustering tools were used to establish space-time patterns in historical events of this phenomenon. While doing so, we assess the limitations of ArcGIS functionalities in handling time. The results shows that ArcGIS have the capabilities of handling time information. Space-time patterns were identified and presented using 2D and 3D visualization techniques. However, more functionalities are still needed in order to efficiently analyse spatio-temporal information, especially in regards to discrete historical temporal patterns. There is still a lack of spatio-temporal query functions. Additionally, more visualization methods are still needed that explicitly accentuate space-time changes without time animations.

Keywords: Spatio-temporal data, Natural Language Processing, Information Extraction, GIS

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

ANNIE	A Newly New Information Extraction
GATE	General Architecture for Text Engineering
GIR	Geographic Information Retrieval
IE	Information Extraction
LHS	Left-hand-side
NE	Named Entities
NER	Named Entity Recognition
NetCDF	Network Common Data Form
NLP	Natural Language Processing
NSA	Namibian Statistics Agency
POS	Part Of Speech
PR	Processing Resources
STT	Spatio-Temporal Data Type
TD	Toponyms Disambiguating
TIR	Temporal Information Retrieval
TR	Toponyms Recognition

CHAPTER 1: INTRODUCTION

1.1 MOTIVATION

Time is indeed a substantial trait of our everyday life. Without time we are not able to discuss past events, changes and future plans. The integration of time functionalities in GIS applications has had favourable research interests in the last years (Nadi & Delavar, 2003). Among others, hardware and software limitations are identified as the drawbacks to implementing temporal GIS applications (Nadi & Delavar, 2003). However, GIS applications are able to efficiently handle the spatial dimension of geospatial data. Consequently, adding functionalities that can accommodate the time aspect of spatial information would be of great significance. The idea is to be able to analyse events in regards to their past, current and future states. The realisation of this concept is only possible when the temporal information is fully integrated and treated as a dimension on its own. I imagine a GIS application with functionalities able to answer questions like “How far is an object from a specific location, at a time a certain event happened?” I would like to believe that this is still possible and this is the motive for this thesis.

On the other hand, there has been research projects on the development of application to handle spatio-temporal data, but more aligned towards big data and multivariate spatio-temporal information. This could be due to the growing trends and developments in mobile technologies and tracking devices that produces enormous amount of data. For example, the development of traffic geo-application by Janković et al. (2017) is one research area that contributed to the design and implementation of applications to handle spatio-temporal in big amounts in GIS. On the contrary, there has been limited research efforts oriented towards the development of GIS applications that facilitate the analysis and manipulation of historical data (Owens, 2007). Due to the fact that historical information describes geography in the past times, to be able to analyse this information we are able not only to understand where we are coming from, but also how we got where we are today and where we are going (Kossatz, 2016). The motivation of this research is to understand the requirements of handling historical spatio-temporal information and to join fellow researchers in the GIScience field in identifying gaps and limitations that hinder the development of temporal GIS applications.

1.2 PROBLEM DEFINITION

When we hear about certain events happening, we tend to be more interested in details situating them in time and space. For example, during September 2017 when hurricane Irma was discovered to hit the south eastern coast the United States of America, the emphasis was more on when and where it was going to hit. This information acquired at the right time can save lives, as people would have time to prepare and take pre-caution measures beforehand. This illustrates how important time and location are. However, dealing with time is not so simple because time is not tangible. Unlike space which can be seen and perceived as physically existing, time cannot be seen (Kossatz, 2016). This makes time hard to scale. It is on this note that we witness the growing research interest in temporal GIS application development to address the time issue.

This research issue has resulted into new developments such as Historical GIS systems that are able to answer questions related to the state of objects in the past and future, such as the HistoGlobe developed by (Kossatz, 2016). When working with historical information there are issues to be addressed too, such as the data sources and information accuracy. For data sources, historical archives contain historical publications and other text documents that are good sources of historical information. This gives us an opportunity to explore the time-geography aspect of published historical events. In this research we embrace the Natural Language Processing techniques to answer questions relating to the extraction of historical information from text publications, which will be used to assess the capabilities of ArcGIS application in modelling, analysing and visualizing time information.

1.3 CASE STUDY BACKGROUND

History tells us that in the western countries discovered an unexplored continent and decided to explore the new horizon. During the 18th century the German settlers arrived in South West Africa currently known as Namibia in hope of settling and taking over the land. After their arrival, they spread across the country and established military stations at different places. Throughout their stay, they employed the local people on their farms to work for them and interacted with the indigenous people in many occasions. The Hereros, who were the dominant tribe in central Namibia at that time, possessed a vast amount of land and cattle. They were in a civil war subsequent to the death of their Paramount Chief. Taking advantage of the disunity, the Germans seized almost a quarter of the land and began to split the Hereros using European settlement schemes (jpeacock, 2016). However, the Hereros revolted in 1904 under Chief Samuel Maharero along with Hendrik Witbooi and struck out against the Germans. According to Klaus (2000) the Hereros besieged Okahandja and Windhoek where the Germans had their military fort around the 12 January 1904, which marks the beginning of the Herero anti-colonial resistance war. The German empire responded to the overwhelming force and pressure from the Hereros to chase them out of the country. They dispatched Lieutenant–General Lothar von Trotha and approximately 15 000 soldiers to Namibia. The plan was to eliminate any threat to their administration in Namibia. Numerous skirmish and battles have erupted at different places throughout the year. On the 11th of August 1904 the Germans met the Hereros at Waterberg plateau where the Hereros got defeated in a fierce battle. This marked the decisive battle for the Herero war (jpeacock, 2016).

For the purpose of this research, we study the time-geography concept of this event as from January to December 1904. The idea is to transform the documented trajectories into representations and to analyse event patterns during the year 1904. This information is used to evaluate the capabilities of ArcGIS to handle historical spatio-temporal data.

1.4 RESEARCH IDENTIFICATION

The performance of a tool or function to carry out an analysis depends on the characteristics of the data being used, such as data type, amount of data and semantics. As mentioned in the previous section, historical spatio-temporal data will be used in this research to assess the capabilities and functionalities of ArcGIS software in storing, manipulating, retrieving, analysing and visualizing spatio-temporal data. The historical spatio-temporal data to be used needs to be extracted from text documents. This research constitutes of two main objectives:

- To develop a framework to automatically extract spatio-temporal and attributive information from text documents.
- To evaluate the effectiveness of tools and functions in ArcGIS in handling time data.

Given the problem definition and objectives above, the next section summarises the research questions to be answered in the thesis.

1.4.1 Research questions

1. What methods are available to recognise and extract spatial and temporal information from text documents?
2. How to extract location event information and produce trajectories from the extracted references?
3. How can historical data be modelled best in regards to:
    a. Temporal vs. spatial data
    b. Precision vs. accuracy of historical information
4. What analysis methods and functions are available for historical spatio-temporal data?
5. What cartographic visualization techniques are suitable to visualize the case study information?

1.4.2 Intended innovations

While joining fellow researchers in identifying the best possible way to handle temporal information, the visualization of such information plays a big role in this study. The novelty of this research is oriented but not limited towards transforming the theoretical history records into a visual interactive story map by using cartographic visualization techniques such as animations. Further innovations are drawn towards the integration of web based technology, especially web mapping technologies.

1.5 MATERIAL AND METHODOLOGIES

1.5.1 Methodologies:

Based on the objectives and research questions outlined in the previous two sections, the methodologies used in this research are as follows:

1. Contextual information extraction
2. Trajectories and location event information extraction
3. Patterns analysis in historical events
4. Time-aware Story Map

1.5.2 Materials:

The primary data sources used in this research are PDF scanned documents from the Namibian Environmental Information System, publications from the Namibian Scientific Society, the Namibian National Archive and online articles. The historical publications used are:

- Let us die fighting (Drechsler, 1966)
- The revolt of the Hereros (Bridgman, 1981)
- Chronology of the Namibian history (Klaus, 2000)
- South West Africa under German rule (Bley, 1971)
- Herero Uprising – Namibia 1 on 1 online article (Namibia 1on1, 2013)

Hardware:

- Personal computer with all peripherals

Software:

- General Architecture for Text Engineering (GATE) developer 8.4
- ArcGIS 10.5 student license
- PostgreSQL 1.6
- Notepad++

Programming languages:

- Python 3.4
- JAPE

CHAPTER 2: LITERATURE REVIEW

2.1 INTRODUCTION

The most important component of this research is gathering the existing concepts and related work to lay the theoretical foundation needed to solve the research questions. There are three key concepts to consider in order to answer our research questions and these are as follows: Spatio-temporal information extraction, modelling time and visualizing time in GIS. The following content discusses the basic theory and existing solutions about these concepts.

2.2 METHODS FOR EXTRACTING SPATIO-TEMPORAL DATA

Text documents have been central to the humanity sciences long before digitization (Gregory, 2002). And as digital text archives become more easily accessible and contain both explicit and implicit spatial temporal information researchers in GIScience have become aware of these new data sources for space-time analysis (Gregory, 2002). Unstructured text documents such as web documents, historical text documents are another sources of spatio-temporal information. Retrieval of such information requires Information Extraction techniques for recognizing and retrieving geographical and temporal entities. There has been a growing research interest in geographic and temporal information extraction from unstructured documents (Strötgen et al., 2010). However, most of these researches focused more on only the extraction of geographical and temporal entities (Bekele, 2014). Therefore, in this research we take a step further to extract attributive information needed to produce trajectories for moving troops.

According to Strötgen et al. (2010), the extraction of spatio-temporal information from documents has been long recognized in the GIScience research field. According to the same author, Spatio-temporal information retrieval is made up of Temporal Information Retrieval (TIR) component and Geographic Information Retrieval (GIR) component both focussing on the extraction and utilization of temporal and geographic expressions from documents. Strötgen et al. (2010) proposed in his research an approach for combining TIR and GIR methods in order to produce spatio-temporal document profiles that allow to explore the geographical content of the document in a chronological order. The idea of combining the two methods is based on the argument that spatial and temporal expressions relating to an event usually co-exist in the document. This approach is relatively suitable for discovery of event location sequences from documents such as travel reports and document related to history.

The roots of geographic information extraction lies with the Named Entity Recognition. (Katz & Schill, 2013) employs an approach that uses Toponyms Recognition (TR) and Toponyms Disambiguating (TD) as main components for geographic information extraction framework. Using the Named Entity Recognition concept, TR focusses on identifying and recognizing the occurring location names in the texts, while TD looks into associating the identified Toponyms with entries stored in a database, mainly used as a gazetteer for cross reference. This introduces a control framework that is aimed at reducing errors that come with natural language. For instance, a sentence may read “Mali is in Africa”, whereby Mali is referring to a person in the document but it could be perceived as referring to the country Mali in Africa. To avoid such confusions, the TD component uses a gazetteer covering the geographic extend of the document to cross reference whether a recognized toponym is contained in the database or not. While this approach seems to be good, it lacks the temporal information retrieval concepts and therefore is not considered in this research.

While most of the research mentioned above concentrates solely on the extraction of geographic information, temporal information extraction is equally important for the scope of this research. (W. Wang, 2014) has used Natural Language Processing concepts for the extraction of spatio-temporal data and semantics extraction for natural hazards from web documents and articles. A similar but more extensive approach is proposed by Bekele (2014) for the extraction of spatio-temporal information from historical gazetteers. The advantage of this approach is that it incorporates Named Entity Recognition concepts 2.2 and rule based method for recognizing patterns in the texts. This also includes spatio-temporal relationship terms such as “*2 days later*”, “*200 km away*”. Unlike the methods in the previously mentioned works, this approach focusses on the extraction of spatial, temporal and attributive information depending on the linguistic and textual information contained in the text documents and this is the approach adopted in this research.

2.3 SPATIO-TEMPORAL DATA MODELS

The fundamental aspect of spatio-temporal modelling is the data model, which determines the manner in which the data is retrieved or presented. A data model in this sense describes the object types, the relations between objects, operation and rules that would be used in order to maintain the database integrity of spatio-temporal database (Nadi & Delavar, 2003). The data model is meant to provide efficient support in handling the data, such as support for spatio-temporal queries. The model depends on a number of issues such as the scale of the data, time density and representations. Spatio-temporal data represents evolution of spatial phenomena in time. The evolution of a phenomenon in GIS can mean change in different aspects that describe the object, such as change in shape and size. The data model should be able to accommodate the changes that comes with time. Just to mention a few, common spatio-temporal data models used in GIS are snapshot model, space-time composite, simple time stamping and event-based model.

2.3.1 Snapshot data model

Snapshot model stores the full state of the object at a point in time t_i in a snapshot. The model does not consider the changes between different time steps. The snapshot model requires a lot of memory, because it stores redundant states of objects where there has been no change. This model is not considered for the scope of this research.

2.3.2 Event-based spatio-temporal data model (ESTDM)

ESTDM model can be compared to a temporal map set that is a raster base instance of a snapshot (Nadi & Delavar, 2003). It is a raster based model that uses a collection of time stamped layers to represent temporal information of an event. Unlike the snapshot model, with this model only changes regarding the previous state are stored.

2.3.3 Simple time-stamp model

Another spatio-temporal data model approach tags each object with a pair of timestamps, one for the time it was created and another for ending time (Nadi & Delavar, 2003). The idea of this model is to facilitate and support simple temporal queries. This model is based on the linear, discrete and absolute time. Only the time that represents the occurrence (valid time) is considered in this model, and is represented as an attribute of the object. Pelekis et al. (2004)

argues that the strength of the simple time-stamp approach lies with the ease to obtain states of objects at certain times, but the disadvantage is that it is not possible to obtain direct information of what happened or why it happened. Pelekis et al. (2004) further noted that this model is strong in answering queries like "What was the state of…" and weak on queries like "What happened".

2.4 SPATIO-TEMPORAL DATABASES

Depending on the type of database management system being used, time information can be stored and managed in different ways. Storing time data facilitates how it can be retrieved, updated, optimized and used for its intended purpose. And for this reason, implementation of a data structure that handles temporal data, or to say temporal database is very important when dealing with temporal information. The major database structures that are commonly used in GIS databases are the Relational and Object-oriented data structures

Relational (RDBMS) - RDBMS models have been widely used in traditional GIS systems for data management. They are built upon the concept of entities and relations. Researchers such as (Wang et al., 2000) have pointed out the inefficiency of relational database structures in handling space and time simultaneously, because of the temporal behaviours and the complex structure that comes with multidimensional temporal data of (x, y, z, time).

Object Oriented (OODBMS) - OODBM refers to database management system that treats and supports the modelling of data as objects and supports classes of objects, inheritance of class properties and methods by sub classes and their objects.

Object – Relational (ORDBMS) - A database management system that combines the concept of relational Data Base Management System (DBMS) while using the object oriented database model is called Object Relational DBMS. It supports objects, classes and inheritance in database schemas as well as query languages such SQL. An example of an ORDBMS is PostgreSQL which is used in this research.

The spatio-temporal data model as discussed in section 2.3 is implemented into a database management system. The idea of a temporal database is to effectively represent attributes of an object as they change in time. Such as having in-built functions that allows queries related to changes in time. A temporal database should be able to capture the time at which an event happened referred to as world or valid time and distinguish it from transactional time which is the time the record is stored in the database. In this research only valid time is relevant.

Effort has been made in research to develop data models that exhibit object oriented concepts for the effectiveness of spatio-temporal data storage, retrieval and management. The idea is to develop data models that handle time as an extra dimension and not as an attribute of a spatial object. For example, PostGIS (PG)- Trajectory data model was developed by Ahmet Kucuk et al.(2016) for PostGIS database with a wide range of functions for storing and manipulating spatio-temporal trajectories.

A similar approach by Frihida et al. (2009) is the Spatio-Temporal Data Type (STT) which is an abstract data type with a set of operations supporting the syntax and semantics of trajectories. The STT data type can be used for the design and implementation of spatio-temporal databases. The data type considers the semantic concepts of a trajectory such as the trip and the static activity by using semantic, spatial and temporal constraints. The biggest challenge of spatial data models in traditional DBMS is the temporal overlaps of objects in space, which make it even harder to model dynamic location events that occurs at same location but at different times. STT data type in its implementation assumes that successive activities do not overlap temporally. It also uses semantic operators to allow querying semantics of the STT such as Activity_Before_Trip: STT × Trip → Activity which returns the activity before the specified trip. It uses temporal operators that express topological relationships in time such as *Time_During_STT* and also spatial operators to model topological relationships in space. Figure 1 illustrates the concept of the STT data type for trajectories.

.

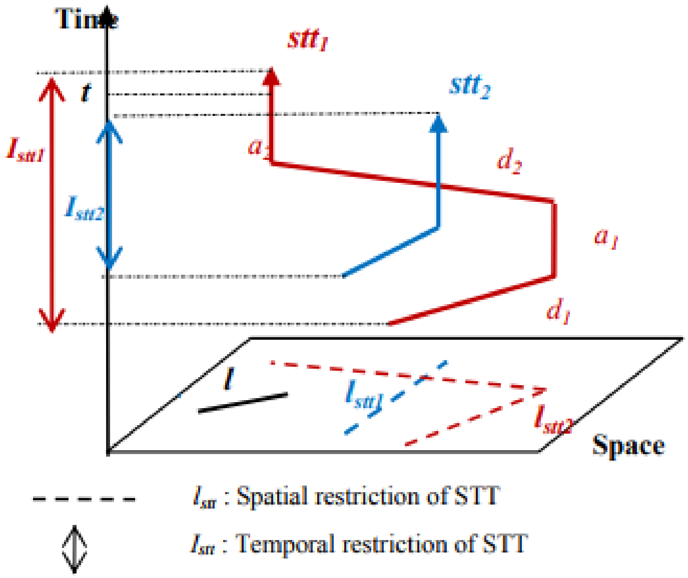


Figure 1. Spatial and temporal restrictions of STT)
Source: *Adapted from Frihida et al. (2009)*

The STT data model can also be implemented in ArcGIS enterprise DBMS through PostgreSQL. It was the initial goal of this research to implement the STT data type but due to software requirements and database configuration setups on the school grounds, it was rather not possible to use and evaluate the effectiveness of this data model in our research.

2.5 VISUALIZING AND ANALYSING SPATIO-TEMPORAL DATA

A historical research normally involves reading a lot of historical data in order to get a clear picture on the historic events. This could be time consuming for some and also it does not guarantee collection of correct information. With the growing technology, current software and tools use the opportunity that comes with modern computer technologies and incorporate the legacy from conventional cartography to represent historical events (Li, 2005). The visualization of historical spatio-temporal data is based on three cartographic depict modes namely: single static map, multiple static maps and animation maps (Kraak 1996).

2.5.2 Cartographic depict modes

2.5.2.1 Single Static Maps

The single static map uses symbols and graphic effects to show change in order to represent an event (Kraak, 1996). Change in this sense is depicted by either colour intensity or symbol size. This is particularly efficient for multivariate spatio-temporal information representation, where no animation is required. The single static map contains all changes for different time periods on one map. Some researchers have thought of the single static map as simple and easy to understand (Li, 2005). However, complex changes with temporal overlaps are difficult to represent in a single static map.

2.5.2.2 Multiple static maps

Multiple series of static maps could be used to represent a temporal sequences of spatial change. The time series concept allows to visualize the change state of an object or event at discrete times, without any overlap. The use of multiple static maps for representation of spatio-temporal phenomena is quite old in GIS (Li, 2005). They are mostly used to illustrate change in spatial growth at different times, such as urban growth maps. The time series maps allow for comparisons of changes spatially and temporally, whereby each map represents one time period. As an illustration, Figure 2 shows the spatial distribution location events during the Herero war of resistance 1904 for three months. Each map represents the location event distributions for a month and comparisons can be clearly made to identify the month with more movements and where more movements where taking place according to the data. However, in cases of long time series this approach may not be useful as it requires large storage capacity for images and it could be hard to manage. Figure 2 demonstrates the use of multiple static maps to illustrate changes in movement for the months of January, February and March 1904.

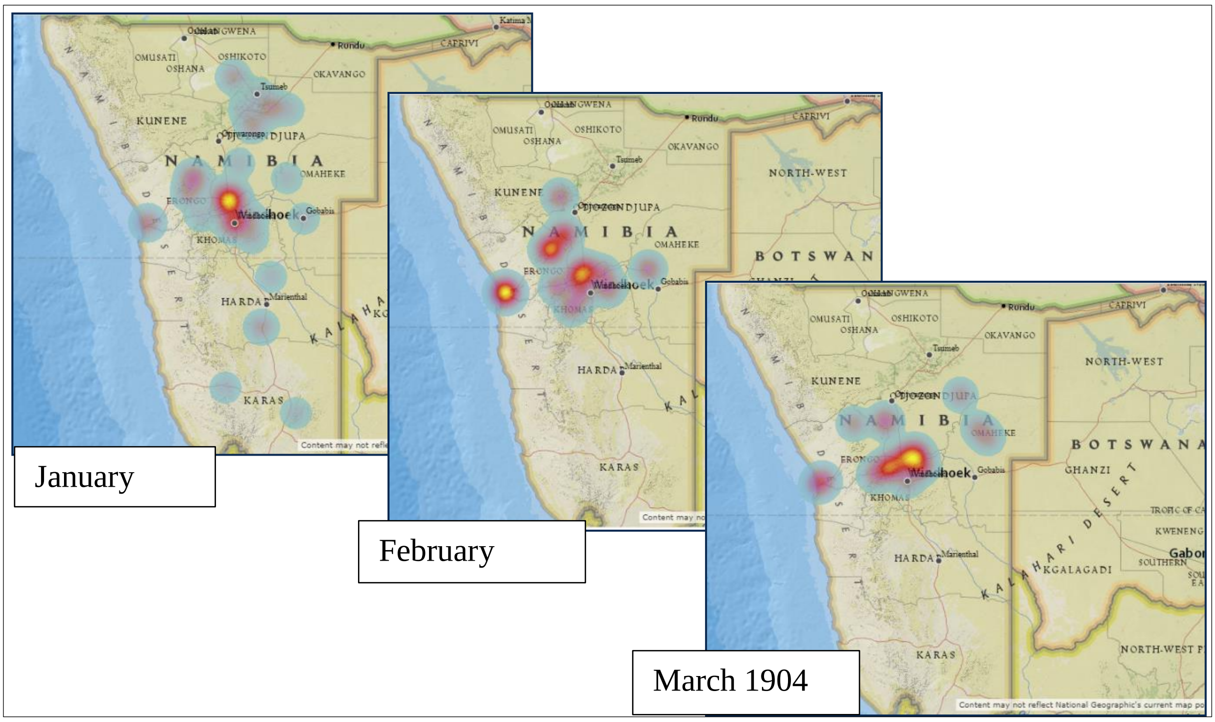


Figure 2. Multiple static maps showing location visit events distribution for January to March 1904

2.5.2.3 Time animations

Lastly, the animation map comes in handy in solving the static map representation issues. Animations stimulate viewers to detect change patterns. GIS applications provides supports for animation map creation. For example, ArcScene allows us to create animations with the animation manager. Additionally, ArcGIS includes time property functionalities that allow to create time animations by configuring time properties on our data. This is particularly relevant for visualization of movement data. Therefore, time the concept of time animation is considered for the visualization of the historical spatio-temporal data in this research.

2.5.3 Space-time visualization models

Space-time cube is becoming a popular method for visualizing spatio-temporal data specifically because it inherently supports three dimensions (x, y for latitude and longitude and z for time). Hägerstrand introduces a space-time model in the sixties which included space-time cube and space time path concepts (Kraak, 2003). Consequent literature in this research area is dominated by space-time cube and space-time path visualizations. For examples work by (Kwan:, 2007) (Li, 2005) (Shaw, 2006). in the space-time cube method by (Kwan:, 2007) a synchronized 3D time layer is overlaid on a 2D to show activities that happened at a location at different times. Figure 3 illustrates the space time cube concept for visualizing women social activities in time and space on a certain day.

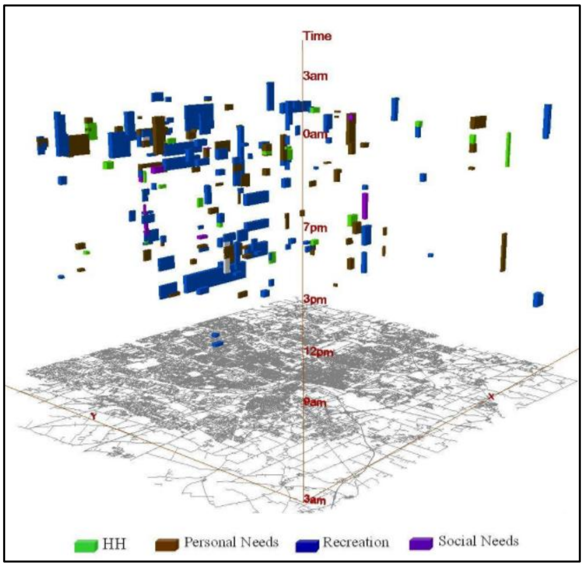


Figure 3. Space-time cube showing women activities in time
Source: *adapted from (Kwan:, 2007)*

In daily life, an individual follows a trajectory through space and time. Space-time paths are similar to space time cube and are used to display trajectories in space and time. The trajectories modelled in a space-time path are influence by constraints namely: capability constrains (such as mode of transport), coupling constrains (such as being at work or sports) and authority constrains (such as accessibility of buildings or parks in space and time) (Li, 2005). This approach allows us to model and understand human movements and interactions with geography in time.

Similarly to (Kwan:, 2007) the space-time path contains a geographical base and time dimension on the z axis as the height. To visualize space paths, this approach is particularly prominent as there are no overlays when the same trajectory line is repeated in time unlike visualizing trajectories in 2D. Figure 4 demonstrates a trajectory path through the city of Enschede on a particular day whereby the vertical line illustrates the time spent at a location or activity to say. The higher the vertical line the more time was spent. This is specifically useful in analysing human activities and understanding the social circles of individuals. The horizontal lines represent the movement from one activity point to another. The length of the horizontal line illustrates the distance between two location points. This method is particularly useful for historical movement data analysis and visualization.

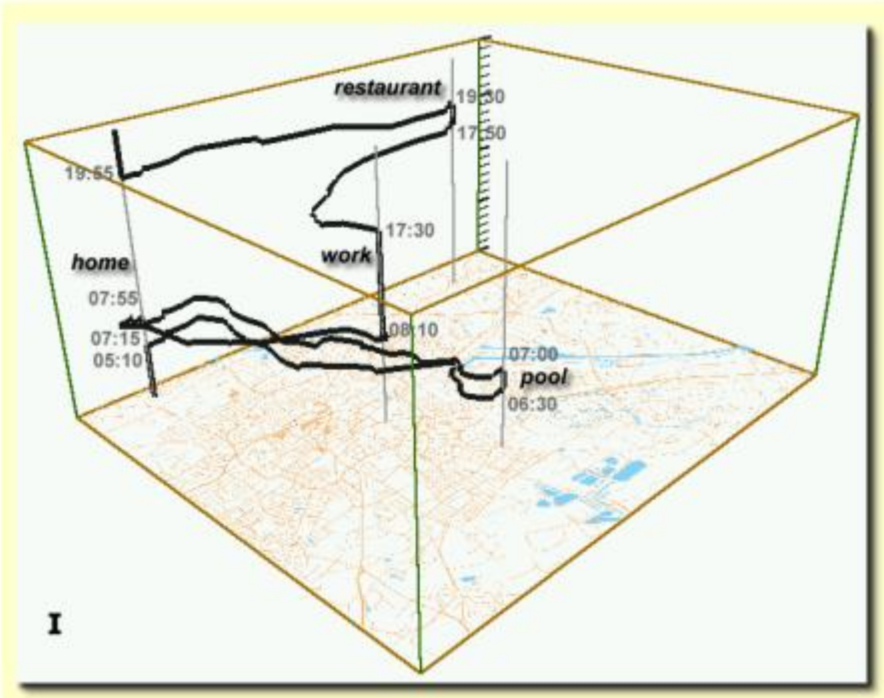


Figure 4.The space-time cube: An example of the a trajectory through Enschede

Source: *Adopted from (Kraak, 2003)*

However, with all these 3D time-geography concepts, researchers still feel the need to improve these methods to offer more capabilities of time-geography analysis and visualization. Shaw (2006) have developed an approach to extend Hägerstrand's time-geography such that it can provide a comprehensive framework to represent and model physical and virtual activities as well as interactions among the objects (Shaw, 2006). The results of this research is a space-time extension to ArcScene that provides interactive functionalities.

CHAPTER 3: EXTRACTION OF SPATIO-TEMPORAL DATA FROM HISTORICAL TEXT DOCUMENTS

3.1 INTRODUCTION

"It's not where you take things from, it's where you take them"

-Jean-Luc Godard

Spatiotemporal Information, as defined in the previous chapter, can be obtained from event happenings on a daily basis. This is because every other event is characterised and identified by the location where it occurred and the time when it happened. Such information can also be found in text documents that are written about events that had happened along with their attributive information. It is on this note that historical publications that describes historical events, are the data sources of spatiotemporal information used in this research.

Researchers in the GIScience field have embraced the Information Extraction (IE) techniques to explore geographical information from text documents and there has been a growing improvements in the field of IE in accommodating geographic information extraction (W. Wang, 2014). Information extraction techniques provide an opportunity to automatically extract spatiotemporal information from text document through Natural Language Processing and Text Engineering methods. In this chapter, we would like to answer the research question on how to abstract spatiotemporal information from historical text documents into spatial representations.

Natural Language Processing (NLP) techniques helps us capture entities with their attributes that can help us answer questions on *what*, *where*, *when* and by *who,* for us to transform the narratives into visual representations in GIS. Rattenbury et al., (2007) defined information extraction as the identification and extraction of instances of a particular class of events or relationships in a natural language text and their transformation into a structure representation, such as a database. Based on this knowledge NLP techniques are adopted in this research to extract contextual information that can describe and answer spatial, temporal and attributive questions pertaining to the movement of the German settlers and the Hereros respectively during the anti-colonial resistance war of 1904 and to explore and build a rich geographic knowledge on this event.

The spatiotemporal information extraction framework adopted in this research comprises of four main components namely: Document pre-processing, Creation of spatiotemporal gazetteers, Trajectory and location event extraction. Creation of spatiotemporal gazetteers and Extraction of contextual information are fully implement within the GATE development framework while Trajectory and location event extraction in Python 3.4 with the aid of psycopg2 adapter that facilitates the communication between the Extensible Markup Language (XML) data source and the PostgreSQL database. The figure below illustrates the framework for automatically extracting spatiotemporal information.

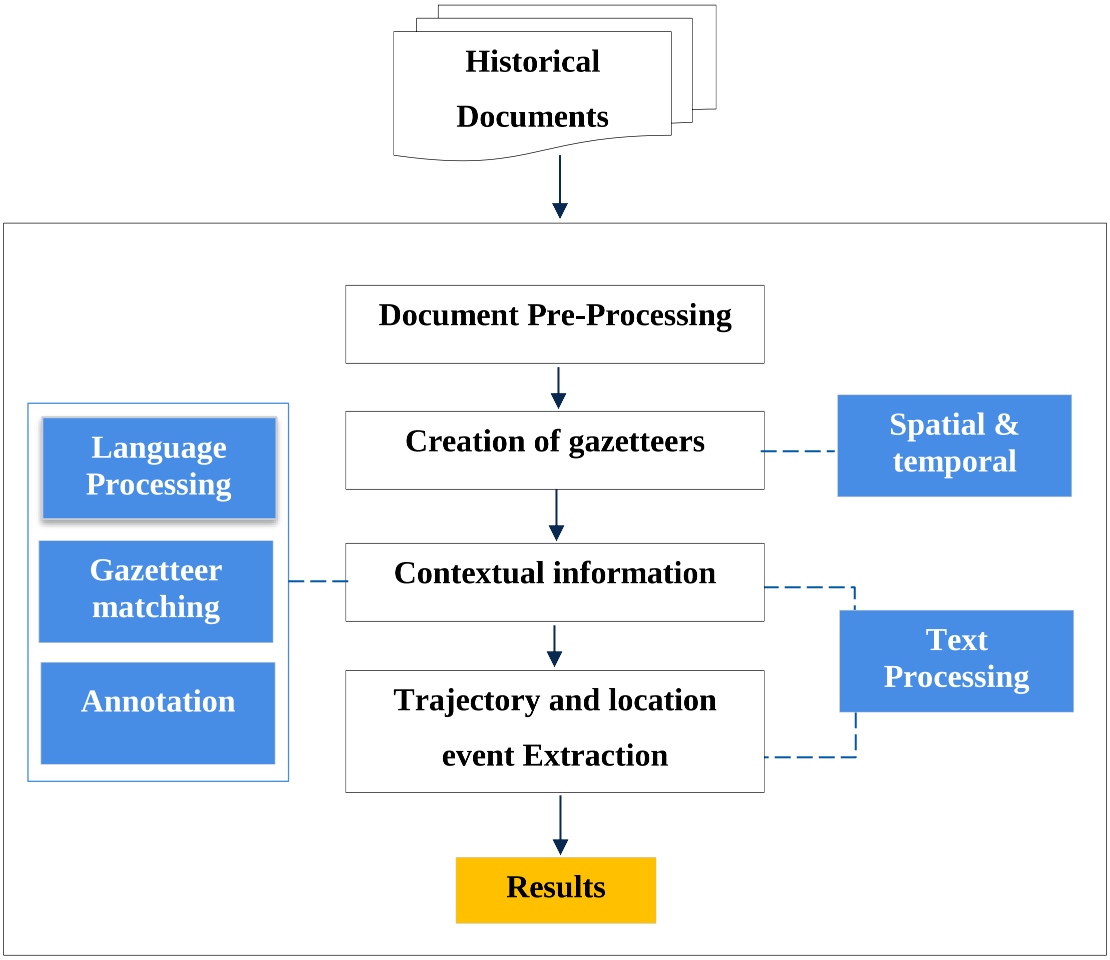


Figure 5. Framework to automatically extract historical spatiotemporal information

The document pre-processing sections handles the raw data conversions and cleaning. For the purpose of the extracting information describing a location visits, the document pre-processing section provides the first transformations of the raw data sources into GATE readable format with a defined structure that would allow for easy annotation and extraction sections to follow.

3.2 DOCUMENT PRE-PROCESSING

The document pre-processing component is the first step of the extraction process as mentioned earlier. The acquired data sources as described in the introduction chapter, are scanned Pdf documents of chapters taken from historical books and online articles. With the proposed approach for information extraction and for GATE development framework used in this research, scanned Pdf images are not recognized and therefore cannot be used in the process unless translated to Microsoft Word, Textfile, Javascript Object Notation (JSON)or XML file. The latter was chosen as the preliminary data source format for GATE annotation framework because of its structure that represents data with descriptive tags for reference and the easy to read tree like structure. The order of the document processing component is depicted in the figure below:

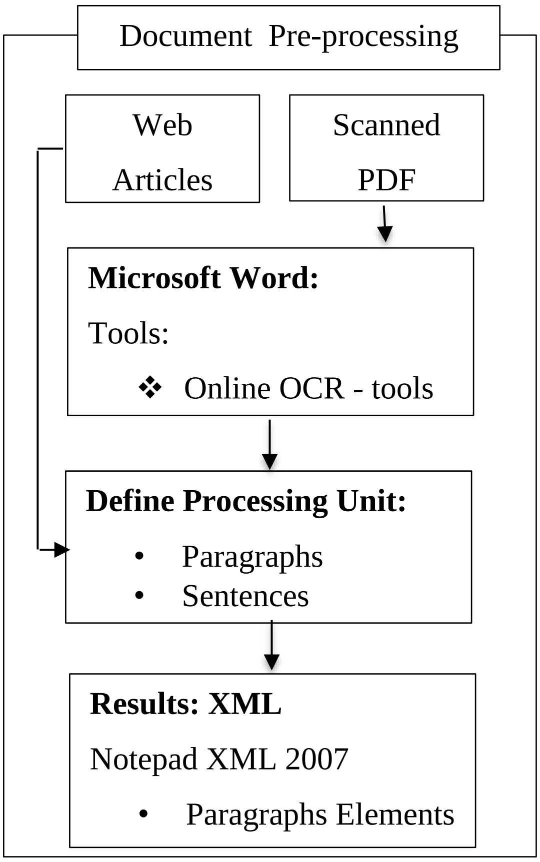


Figure 6. Document pre-processing framework

The scanned Pdf images and online article contents are all transformed in to XML tree like structure with paragraph elements and sentence elements using Notepad XML 2007. The idea is to keep the order of the events as described in the documents and also to define a processing unit which represents an event description. A sentence in this case is used as a reasoning unit for an event as it contains most information needed for a location event description. It appears that temporal, spatial information and attributive terms relating to one location event are most likely to be found within one sentence and the sentences following after that would be containing temporal and spatial terms for the next location event. This all depends on different authors and the way they express and write. For example, the figure below shows the paragraph describing location events that are related spatially and temporally. A paragraph in this sense contains events that happened within the same temporal and spatial boundary. The pattern in which location visit events are described in the document is very important in this section to distinguish events from each other. Figure 3 below shows a paragraph from the Herero Uprising article online, which depicts how a paragraph contains location visit event information within sentences. The second sentence in the figure is a continuation on the first event but literally a different location visit event. Therefore a sentence is treated as a location visit unit in this research.

> 6 **25 February** - Otjihaenamaparero: Major von Estorff in command of two companies of Schutztruppe along with one company of Marines who had recently arrived in the country moved westward from Okahandja to seek out the Hereros who had moved away from the broken siege at Omaruru. It was at Otjihaenamaparero that they met and fought with about 1000 Herero warriors. Fierce fighting ensued for about 12 hours. The Germans lost 12 men, and estimated Herero losses were put at 50 men.

Figure 7. Screenshot of the paragraph in the article

Source: *Screenshot from (Namibia 1on1, 2013)*

In addition, the paragraph contains a bolded date at the beginning which implies that the events described within the paragraph happened either on that date or close dates if no other temporal expression is found within the sentences. It is for this reason that the date at the beginning of the paragraph is captured in the XML as a date attribute for the paragraph element. This is to be used as a reference date later for the location visit events from the same paragraph that do not have temporal expressions when combining the extracted spatial and temporal references together. The results of data pre-processing section is an XML with a structure that is ready for annotation in GATE framework. Figure 4 shows the XML document from the document processing component.

```xml
<paragraph id="5000" date="10.08">
  <sentence id="5001">10.08. Von Trotha plans the final battle from his headquarters at Ombuatjipiro.</sentence>
  <sentence id="5002">He put his plans in his own words: "My initial plan for the operation, which I always adhered to, was to encirc
  <sentence id="5003">The German troops have the following positions on this day: Unit Von Estorff at Okomiparum.</sentence>
  <sentence id="5004">Unit Von der Heyde at a position 15 km north east of Hamakari (Ohamakari).</sentence>
  <sentence id="5005">Unit Mueller at Ombuatjipiro.</sentence>
  <sentence id="5006">Unit Deimling at Okateitei.</sentence>
  <sentence id="5007">Unit Von Fiedler at the Osondjache Mountain and Unit Volkmann near Otjenga.</sentence>
</paragraph>
<paragraph id="5100" date="11.08">
  <sentence id="5101">11.08. The Waterberg battle begins.</sentence>
  <sentence id="5102">The fighting takes place mainly at the areas southeast of the Waterberg (Klein Hamakari and Hamakari (Ohamakari
  <sentence id="5103">There are great losses on both sides.</sentence>
  <sentence id="5104">The heaviest fighting occurs at the Hamakari waterhole.</sentence>
  <sentence id="5105">The main German section under Von Trotha advances from Ombuatjipiro to Hamakari.</sentence>
  <sentence id="5106">Berthold von Deimling proceeds from Omuveroume.</sentence>
  <sentence id="5107">Von der Heyde attacks from Okakarara, east of Hamakari.</sentence>
  <sentence id="5108">At Otjosongombe Von Estorff starts firing on Ovaherero, and defeats them early on 12.08.</sentence>
  <sentence id="5109">All other advances planned by the Germans fail on this day.</sentence>
  <sentence id="5110">Von Deimling does not succeed in realising Von Trotha's plan to trap and defeat the Ovaherero.</sentence>
  <sentence id="5111">An official report later announces: "The bold enterprise shows up in the most brilliant light the ruthless ener
  <sentence id="5112">No pains, no sacrifices were spared in eliminating the last remnants of enemy resistance.</sentence>
  <sentence id="5113">Like a wounded beast the enemy was tracked down from one waterhole to the next, until finally he became a victi
  <sentence id="5114">The arid Omaheke was to complete what the German army had begun: the extermination of the Herero nation."</sent
  <sentence id="5115">Major Stuhlmann describes in his diary for this day a scene from the battle of Hamakari where he reflects on th
  <sentence id="5116">Many dead Ovaherero soldiers are buried by the Germans on Hamakari (Ongwero).</sentence>
</paragraph>
<paragraph id="5200" date="12.08">
  <sentence id="5201">12.08. Von Deimling advances to Hamakari, and this is the last straw for the Ovaherero who start fleeing in a s
</paragraph>
<paragraph id="5300" date="13.08">
  <sentence id="5301">13.08. Berthold von Deimling and Karl Ludwig von Mühlenfels set off in hot pursuit of the main group of Ovahere
  <sentence id="5302">A one-day delay gives Samuel Maharero a lead and saves his life because the Germans are unable to catch up.</se
  <sentence id="5303">But a tragic scene unfolds: a nation flees without food or water.</sentence>
```


Figure 8. XML from document pre-processing section

3.3 CREATION OF GAZETTEERS

A gazetteer is a geographic index or dictionary as defined by the Oxford dictionary. The gazetteer provides a list of place names and their alternative names which is used to find matching patterns in the text documents. In this section, spatial and temporal gazetteers were created to help in entity recognition and extraction from the source documents.

Named Entities (NE) in this section refers to main elements in texts belonging to predefined categories such as persons, organisations, location, date expressions, etc. While Named Entity Recognition (NER) describes the task of identifying these named entities in the texts documents. With GATE annotation framework there are two ways to identify entities in an input text: 1. By matching entities stored in named entity lists in the gazetteer against the input texts in the document. 2. By matching the patterns of entities with JAPE grammar rule which is explained more in section 3.4. All entities that do not require pattern matching such as months, person's names, direction indicators and place names are recognized by gazetteer matching and they need Named entity lists to be created. The main gazetteers for this research are spatial for place names and spatial expressions and temporal gazetteers for temporal expressions.

3.3.1 Spatial Gazetteer creation

Spatial gazetteers plays an important role in Geographic Information Retrieval. One important point to consider when building an information extraction application using NLP and specifically GATE software, is the ability of the gazetteers to capture the domain specific information and also the geographic extend of their spatial gazetteer. For the purpose of this research, the gazetteer to be used ought to recognize the traditional location names for Namibian places. However, the A Newly New Information Extraction System (ANNIE) gazetteer is UK based and it lacks most place names found in the historical data sources. Most place names mentioned in the documents for the Anti-colonial war in Namibia are historical names which have been renamed after the Namibian Independence in 1991. Therefore, these historical place names and their alternative names had to be collected and added into the gazetteer for these place names to be recognized. The local place names have been gathered from the Namibian Statistics Agency (NSA), Namibia Environmental Information Service online and from Geonames gazetteer online.

The place names are stored in list files (.lst), which are plain text entries with one entry per line. Each list file contains entities of the same kind. All the list files are defined and accessed through the “list.def” index file. The definition of each file contains the file name, major type, minor type, and language and annotation type in five columns separated by colons. The columns are defined as below:

Column 1: Defines the list file name and the extension.

Column 2: Defines the primary identifier to be used when an entity in the input text is found in the list file. This is called “MajorType” and if not defined it will be assigned to lookup type. For example, the major type for the word “Dresden” is Location. The major type definition is compulsory.

Column 3: Defines the secondary identifier called “MinorType”. The minor type comes always with the primary identifier. For instance, the minor type for “Dresden” is “City”.

Column 4: Defines the language of the input text being processed. This is optional and can be omitted.

Column 5: Defines the annotation type used when an entity is found in the list file. For example, the major type for Dresden is Location and the minor type is City, the annotation type would be LocationCity. When “Dresden” text is found in the input text will be tagged as “LocationCity” type.

The figure below shows the definition file:

```
person_male_lower_ambig.lst:person_first:male
person_female_lower_ambig.lst:person_first:female
religious_adj.lst:religious_adj
nam_towns.lst:location:town
nam_placenames.lst:location:village
nam_region.lst:location:region
nam_personnames.lst:person_full:person_male
event_words.lst:event:other
direction_indicators.lst:direction:other
spatial_indicators.lst:spatial_indicator:other
month.lst:date:month
nam_group.lst:group:person
```

Figure 9. List definition file

3.3.2 Temporal gazetteer creation

Extracting temporal references from text documents in GATE can limit the possible amount of information to be extracted because not all temporal expressions are captured in the gazetteer. Not all temporal terms occur in a definite and usual date formats such as DD-MM-YYYY. There are a lot of possible temporal expressions that should to be identified and extracted in order to obtain full temporal coverage of the document. For that we had to extend the GATE’s temporal gazetteer to identify most temporal references used in the publications, for example *“11.09”*, *“January 13, 1904”*. A gazetteer for the different date formats is created using JAPE grammar rule and stored as a JAPE transducer called dates.jape file.

The temporal entity gazetteer constitutes of the following date formats:

Table 1: Temporal entity gazetteer

No.	Entity	Pattern
1.	Date	June 1904
2.	Date	June 13
3.	Date	June 13, 1904
4.	Date	13 June
5.	Date	13 June 1904
6.	Date	11.06.
7.	Date	11.06.1904

The spatial and temporal gazetteers are input the next component, the contextual information extraction component. The Contextual Information Extraction component take in the gazetteers and the XML documents from the Document pre-processing component.

The main concept for the extraction of contextual information is to parse the sentences in the XML which represents location visit events, match the gazetteer against the input files, annotate and tag the names of the people, the spatial information, and the temporal information and within the sentence as explained in the next section.

3.4 CONTEXTUAL INFORMATION EXTRACTION

Contextual Information Extraction is the third phase of the spatio-temporal data extraction process. This phase contains two main extraction pipelines namely: Entity Extraction pipeline and Spatio-temporal relationship Extraction pipeline. Both the pipeline are developed in GATE development framework using the ANNIE gazetteer and JAPE Grammar rules. Entity Extraction pipeline is run first and the Spatio-temporal Relationship pipeline is run second. The order is very crucial because the Spatio-temporal relationship pipeline uses the annotation results from the Entity extraction pipeline as inputs. Therefore, it can only be run after the Entity Extraction pipeline is run.

The contextual information extraction process takes the XML document from the document pre-processing phase as an input and returns another XML document with annotations of the named entities as specified in the gazetteers. These annotations will be combined and pushed to the database at a later stage. Before running the pipelines, the Named Entities must be correctly specified and the JAPE transducer for named entities must be tested to make sure that they are recognizing the patterns.

Figure 10 illustrates the components of the Contextual Information Extraction process.

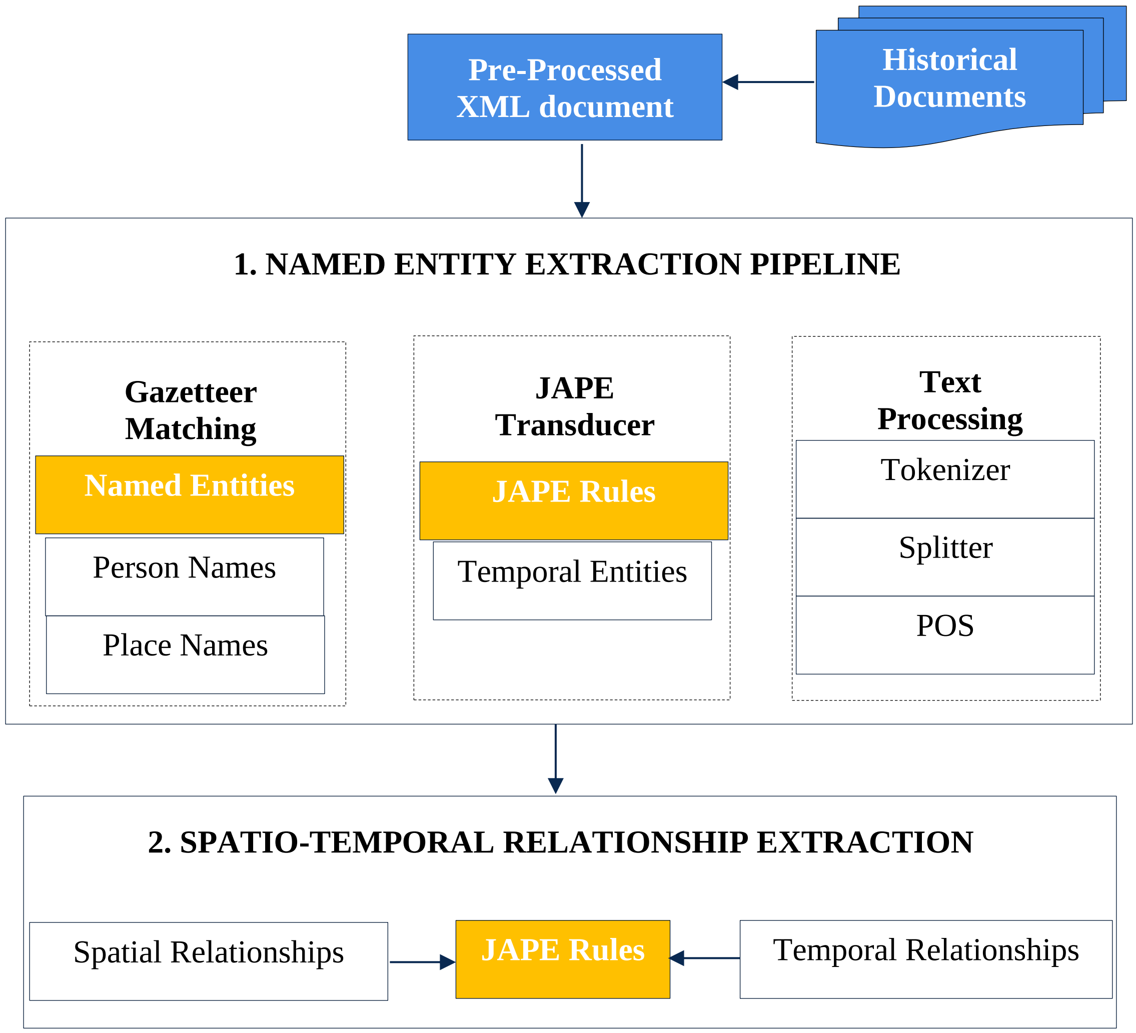


Figure 10. Contextual Information Extraction Framework

3.4.1 Named Entity Extraction pipeline

The Named Entity Extraction pipeline identifies and annotates the needed elements in the XML documents. The named entities are place names, date expressions and people's names are recognized from the texts and annotated. The Entity Extraction pipeline contains 3 processing sections namely: Gazetteer processing, Text Processing and Java Annotated Pattern Engine (JAPE) transducer processing. The Text Processing is done first to prepare the texts contents for further processing. Subsequently, the gazetteer matching and the JAPE transducers follow. Named entity recognition is done either by the gazetteer processing or the JAPE transducer, depending on the entity being annotated. In this section, all named entities are to be annotated and assigned to an annotation class. After this stage, the persons names, location names, dates are annotated and are ready for the spatiotemporal relationship extraction pipeline.

3.4.1.1 Components of the Text Processing Machine

The Text processing machine acts as a temporal annotator, in this case it does text processing using the Processing Resources (PR) available in GATE ANNIE. The reason for having the text processing machine before the NE extractions resources is that it recognizes the sentences, words, punctuations and language resources needed in the next stages. The processing resources used in this section are:

- *English Tokenizer*
  The Tokenizer splits the contents into Tokens types such as words, punctuations, numbers and space tokens which will be useful for JAPE as the grammar rule uses the token types to look up for entities.
- *Sentence Splitter*
  This module is responsible for segmenting texts into sentences. Once the sentences have been recognized and defined, other processing resources such as Part Of Speech (POS) can make use of the texts in the sentences to categorize them accordingly. For this reason, the sentence splitter is run first before the POS. The sentence splitter uses a list of gazetteer list of abbreviations that help distinguish sentence-marking full stops from other kinds.
- *Part Of Speech Tagger (POS)*
  The Part of Speech tagger is the last text processing tool that assigns POS tags to each annotation token, the tags mostly indicate a linguistic category which a token (e.g. word) belongs to, for instance NNP for Proper Nouns in singular form. (GATE reference) These tags are mostly used by JAPE transducer when annotating entities and this is another reason why the JAPE transducer is run after the Text Processing tools.

3.4.1.2 Entity Recognition by Gazetteer matching

As mentioned before, named entity recognition can be done either by gazetteer matching or by JAPE transducer. Once again, the named entities extracted for this research are place names, people's names and temporal entities. The gazetteer recognizes and annotates the entities in the document by matching them against the reference datasets in the gazetteer .lst files created in the section before. The gazetteer annotator takes the words in a sentence and matches them against the words in all the lists in the gazetteer, if there is a match found the word gets an annotation class tag as defined in the list.def file.

3.4.1.3 Entity Recognition by JAPE Transducer

JAPE is a finite state transducer that operates over annotations based on regular expressions (Thakker, Osman, & Lakin, 2009). It is useful for semantic extraction and it used in this section because of its ability for pattern matching. A JAPE grammar consists of a set of phases, each of which consist of a pattern or rules. For instance, the temporal transducer used to match the temporal entities in this research contains a total of seven date patterns within one transducer, of which each pattern has its own rule describing that pattern. For named entity recognition in this section, two transducers are used: the temporal transducer, special names transducer. JAPE uses a left-hand-side (LHS) and (RHS) right-hand-side rule. Whereby the LHS consists of the annotation pattern description while the RHS consists of the annotation manipulation statements. The logic is, once the pattern on the LHS is matched, do what the RHS says. Figure 11 shows an example of the JAPE grammar rule to match a temporal pattern in the document:

3.4.1.3.1 Temporal Entity Transducer

Different authors uses different patterns for temporal expressions. Seven different date formats were identified in the different documents and JAPE grammar rules were developed for each date format. All the seven data formats are written in one .jape file which is called into the main.japefile in GATE annotation platform. The temporal transducer uses the annotation classes from the gazetteer processing machine and the text processing. The annotation classes for the date patterns are as follows:

1. *MonthYear*: If a month is followed by a four digit number, it is annotated as *MonthYear*, such as *July 1904*.
2. *MonthDate*: If a month is followed by a one digit or a two digit number, it is annotated and assigned to annotation class *MonthDate*, such as *July 13.*

3. *MonthDateYear*: If a month is followed by a one digit or two digit number which is followed by a ‘,’ and a four digit number, it is assigned to annotation class *MonthDateYear*, such as *July 13, 1904*.
4. *DateMonth*: If a one digit or two digit number is followed by a month, it is annotated and assigned to annotation class *DateMonth*, such as *13 July*.
5. *DateMonthYear*: If a one digit or two digit number is followed by a month followed by four digit number, assign it to class *DateMonthYear*, for instance *13 July 1904*.
6. *NumberDate*: If a two digit number is followed by a ’,’ or ‘-’ or ‘.’, and followed by a two a digit number, it is assigned to annotation class *NumberDate*, for instance *13.07*.
7. *NumberDateYear*: If a one or two digit number is followed by a ’,’ or ‘-’ or ‘.’, it is assigned to *NumberDateYear* annotation class such as *13.07.1904*.

```
Phase: dateExpressions
Input: Token Lookup SpaceToken
Options: control = appelt

////////////////////////////////////Macros
//Initialization of regular expressions
Macro: DAY_ONE
({Token.kind == number,Token.category==CD, Token.length == "1"})

Macro: DAY_TWO
({Token.kind == number,Token.category==CD, Token.length == "2"})

Macro: YEAR
({Token.kind == number,Token.category==CD, Token.length == "4"})

Macro: MONTH
({Lookup.minorType=="month"})


//////////////Date pattern 3
//For date format June 13, 1904
Rule: monthdateyear
Priority: 50
(
     (MONTH)
     ({SpaceToken})
     (DAY_ONE|DAY_TWO)
     ({Token.string == ","}|{Token.string == "."} |{Token.string == "-"})
     ({SpaceToken})?
     (YEAR)
)
:monthdateyear
-->
     :monthdateyear.NumberDate= {rule = "monthdateyear"}
```

Figure 11. JAPE grammar rule example

The results of the Named Entity Extraction pipeline are annotated texts assigned to their annotation classes. Information such as “12km from the river” and “2 days later” are spatiotemporal relations that expresses the temporal period and spatial location of an event in a different way. Equally important, spatio-temporal relationship terms allow the extraction of full spatial and temporal coverage of the story. Natural Language Processing provides the techniques to abstract implicit patterns that can help us answer questions, when no answer is provided. The next section discusses the recognition of spatial and temporal relationships terms in this research.

3.4.2 Spatiotemporal relationship extraction pipeline

The essential function of a language is the ability to express spatial relationships between objects and their relative location in space (Kordjamshidi et al., 2011). Such relationship expressions may also be useful as they give an approximation on where or when something happened. The spatiotemporal relationship extraction pipeline consists of two transducers, the (i) spatial relationship transducer and the (ii) temporal relationship transducer. As mentioned earlier, the spatiotemporal relationship extraction pipeline should be headed by the entity extractions because it uses the annotation classes of the entity extraction pipeline as its inputs.

3.4.2.1 Spatial Relationship Transducer

The spatial relationships transducer takes the annotated location entities from the entity extraction component and annotate the relationship between the location entities. The spatial relationship transducer contains 14 rules for the spatial expression pattern found in the documents. See Appendix A for the spatial transducer with all the rules. Spatial Relationship rules in this research are as follow:

- *Rule 1*: Between location and location e.g. Between Windhoek and Karibib.
- *Rule2*: Directional indicator followed by the locations e.g. North of Windhoek.
- *Rule3*: Distance followed by a direction indicator, followed by a location e.g. 20km north of Omangeti.
- *Rule4*: Spatial term followed by location, followed by another location and another location e.g. Areas of Omangeti, Windhoek and Rehoboth.

- *Rule5*: Directional indicator followed by "of", "the" and location e.g. South of the Great Waterberg
- *Rule6*: "from" followed by location, followed by "via" location, followed by location, followed by "and", followed by location e.g. from Otjimanangombe via Epata, Otjosondu and Osondema to Otjituuo
- *Rule7:* "about" followed by distance e.g about 25 miles.

3.4.2.2 Temporal Relationship Transducer

The temporal relationship transducer works similar to the spatial relationship transducer. Here, expressions such as *Between June and July* or "*the day of the battle"* are extracted and assigned to temporalrelation's annotation class.

The figures below summarize the annotated entities and relationships by the gazetteers or by JAPE transducer respectively.

Table 2: Entities extracted by JAPE Transducer

No.	**Entity**	**Pattern**	**Annotation Class**
1.	Date	June 1904	MonthYear
2.	Date	June 13	MonthDate
3.	Date	June 13, 1904	MonthDateYear
4.	Date	13 June	DateMonth
5.	Date	13 June 1904	DateMonthYear
6.	Date	11.06.	NumberDate
7.	Date	11.06.1904	NumberDateYear
9.	SpecialNames	Great Waterberg, Little Waterberg	Location
10.	Unit Names	Unit Erstoff, Unit von Trotha	Units

Table 3: Relationships extracted by JAPE

No.	Entity	Pattern	Annotation Class
1.	Spatial Relationship	20km southwest of the Great Waterberg	SpatialRelation
2.	Temporal Relationship	2 days later, the day of the battle	TemporalRelation
3.	Temporal Relationship	The day of the battle	TemporalRelation
4.	Temporal Relationship	between June and July	TemporalRelation

The annotations in GATE application is as follows:

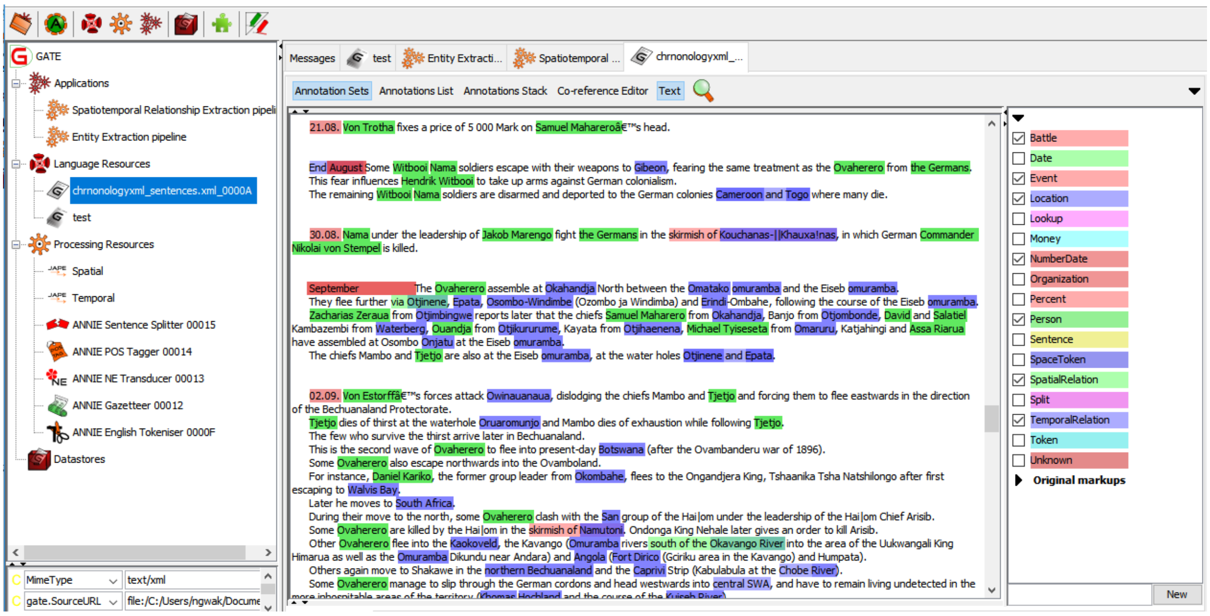


Figure 12: Entity recognition in GATE

Each annotation class is represented by a colour, making it easy to read. The annotation list on the left allow contains classes of entities recognized in the document. GATE also allows to annotate on the fly when there is an entity that is assigned to a wrong class. While on the right is the list of the Processing resources, the corpus containing the text documents and the pipelines.

The annotation results area as follows:

```
ml version="1.0" encoding="UTF-8"?>
cument xmlns:gate="http://www.gate.ac.uk" gate:gateId="0" gate:annotMaxId="28690" title="The Resistance Struggle culminates in genocide: 1904-1906">
<paragraph gate:gateId="1" date="11.01" id="100">
  <sentence gate:gateId="2" id="101">
    <NumberDate gate:gateId="28273" rule="numberdate">11.01.</NumberDate>
    <Person gate:gateId="28443" firstName="Samuel" rule="PersonFull" ruleFinal="PersonFinal" gender="male" surname="Maharero" kind="fullName">Samuel Maharero
    orders all
    <Person gate:gateId="28444" rule="GazPerson" ruleFinal="PersonFinal" surname="Ovaherero" kind="fullName">Ovaherero</Person>
    chiefs to take up arms against
    <Person gate:gateId="28445" firstName="the" rule="GazPerson" ruleFinal="PersonFinal" surname="Germans" kind="fullName">the Germans</Person>
    .
  </sentence>
  <sentence gate:gateId="3" id="102">He orders them to &amp;quot;refrain from touching missionaries, English, Basters, Berg-Damaras, Namas and Boers&amp;quot;
  <sentence gate:gateId="4" id="103">There are doubts concerning the date of this order.</sentence>
  <sentence gate:gateId="5" id="104">
    It is possible that
    <Person gate:gateId="28446" rule="GazPerson" ruleFinal="PersonFinal" surname="Maharero" kind="fullName">Maharero</Person>
    wrote this letter after the outbreak of the war (around
    <NumberDate gate:gateId="28274" rule="numberdate">20.01.</NumberDate>
    ), after the first shots were fired in
    <Location gate:gateId="28447" rule="InLoc1" ruleFinal="LocFinal" locType="town" kind="locName">Okahandja</Location>
    , where it is not clear at all, who actually fired these first shots (Missionary Diehl reports that only
    <Person gate:gateId="28448" firstName="the" rule="GazPerson" ruleFinal="PersonFinal" surname="Germans" kind="fullName">the Germans</Person>
    fired on his house, not the
    <Person gate:gateId="28449" rule="GazPerson" ruleFinal="PersonFinal" surname="Ovaherero" kind="fullName">Ovaherero</Person>
    ).
  </sentence>
  <sentence gate:gateId="6" id="105">
    <Person gate:gateId="28450" firstName="Samuel" rule="PersonFull" ruleFinal="PersonFinal" gender="male" surname="Maharero" kind="fullName">Samuel Maharero
    tries to involve the Basters, under
    <Person gate:gateId="28451" firstName="Hermanus" rule="GazPerson" ruleFinal="PersonFinal" surname="van Wyk" kind="fullName">Hermanus van Wyk</Person>
    and
    <Person gate:gateId="28452" firstName="Hendrik" rule="GazPerson" ruleFinal="PersonFinal" surname="Witbooi" kind="fullName">Hendrik Witbooi</Person>
    , in the struggle. The two letters
```


Figure 13: Annotation results in XML

3.5 TRAJECTORY AND LOCATION EVENTS EXTRACTION

The fourth component is the extraction process which involves producing location visit events from parsing and combining annotated spatial and temporal terms together with the attributive information such as the person's names. A person's name is the name of the commander of a troop or unit. The location visit event is described by the location name, date and the troop that was at that location. The approach of automatically extracting spatiotemporal information from the text document combines the annotated elements using the order they appear in the document. The extracted location visit descriptions are automatically written to PostgreSQL. This approach is designed and developed in Python 3.4. The main requirement for a location event is the troop's name. This indicates which troop was presents at the location, which is of substantial importance in this research. It is for this reason we are able to produce individual troop trajectories from the location visit events.

```
Input: XML Document D, Paragraph P, Sentence E
Results: combine[T, S, N]
where T= Temporal term, S = Spatial term, N = personNames

Begin:
        Parse D,
        For each Paragraph P in D do:
                Get paragraph date as Pd
        For each Sentence E in P do:
                If only S and N then
                        assign Pd as T
                        combine (T, S, N)
                If only one T, one S and N then
                        combine(T, S, N)
                If multiple T and one S then
                        assign S to each T , combine(T1, S, N), combine(T2, S, N)....
                If  multiple S and one T then
                        assign T to each S, combine(T, S1, N), combine(T, S2, N).....
                If multiple S and multiple T and one N then:
                        if S == T then
                                combine(T1, S1, N), combine(T2, S2, N)....
                If multiple T, multiple S and multiple N then
                        if T == S == N then
                                combine(T1, S1, N1), combine(T2, S2, N2)....
        Else
        Jump to next sentence
        Return combine(T, S, N)
End
```

Figure 14: Algorithm to extract location visit events

The extraction approach considers the textual arrangement in the sentences and assigning appropriate location, date and person's names to the appropriate location visit. As mentioned earlier, the algorithm developed takes the sentence as the processing unit for a location visit event and extract the spatial and temporal information jointly using the developed rules. The temporal term is assigned from the location name or temporal relationship term. The same applies to the spatial term. However, not every sentence with a person's name would have the spatial and temporal information and let alone the order in which the spatial and temporal expressions appears in the sentences. The algorithm above is set to handle the cases where multiple persons, multiple temporal and spatial information is found in a sentence. According to (Stewart Hornsby and Wang 2010) there are 5 cases that can arise when extracting spatiotemporal information from a sentence. In our case, the following possibilities are found:

1. Only spatial term and person names are present
2. One spatial term and one temporal term

3. Multiple temporal terms and one spatial term
4. Multiple spatial terms and one temporal term
5. Multiple spatial terms and multiple temporal terms.

Only Spatial information is found

There were a lot of cases were only persons and spatial information are found in a sentence, for example, "*Tjetjo's group has been hiding in Okahandja, trying to escape the Germans.*" Tjetjo is a Herero Chief and Okahandja is a location name. The algorithm check the whole sentence if there is a temporal term and if not, it checks if there is a paragraph date and assigns it as a temporal reference.

When only one spatial term and one temporal term

Some sentences would have only one spatial term and one temporal term available. In this case the spatial term and the location term are combined into the location event description together with the persons names found in the sentence. For instance, "*On the 06.01 Kurt Streitwolf reports on a meeting with Traugott Tjetjo in the Gobabis district"* where *06.01* is a temporal term, *Kurt Streitwolf* and *Traugott Tjetjo* are person's names and Gobabis a spatial term. In such a case where only one spatial term, one temporal term but multiple persons names, the algorithm takes all the persons, the spatial term and temporal term and combine them into a location event description.

When there are multiple temporal terms and one spatial term

Sentences with multiple temporal terms and one spatial term can be a little tricky. In this case, the spatial term is assigned to each temporal term respectively. This will then constitute multiple location events on different dates. For example, "*Leutwein have been involved in skirmish with Ovahereros in Omaruru on the 12.01. and on the 17.01."* In this example, Leutwein is a German commander, Omaruru is the spatial term while 12.01 and 17.01 are the temporal terms. The algorithm creates two location visit event description for each date e.g. *[Leutwein, Omaruru, 12.01]* and *[Leutwein, Omaruru, 17.01]*.

When there are multiple spatial terms and one temporal term

It is also possible for one temporal term to exist with multiple spatial terms in a sentence such as, "*Von Glasenapp's unit remains defensive for the time being and is allowed to march to*

Otjihangwe and later to Otjihaenena arriving on 24.04." In this case, Von Glasenapp is a German unit, Otjihangwe and Otjihaenena are location names and 24.04 is the date. It is not clear as to when exactly they arrived in Otjihangwe but the sentence says they arrived on the 24.04 in Otjihaenena. For this case, it is assumed that they arrived in both locations on the 24.04 and two location visits will be extracted for Otjihangwe and Otjihaenena.

When there are multiple temporal terms and multiple spatial terms

There are cases where multiple temporal terms and multiple spatial terms exist in a sentence. By looking into the data sources, in most cases this appears when describing the positions of different force units at different dates. And for this reason, it is assumed that the terms close to each other are of the same location visit, for example, The uprising is triggered off at different times: Okahandja: 12.01 by Chief Zeraua; Omaruru: 17.01 by Maharero and Otjimbingwe: 23.01 by Riraua. In this example, Okahandja, Otjimbingwe and Omaruru are spatial terms and the temporal terms are 12.01, 17.01 and 23.01. In cases like this where the spatial terms and temporal terms are equal to the number person names in the sentence, the algorithm assigns the first temporal term, fist spatial term and first person name to one location visit description and the second temporal term, second spatial term and second person to new location visit, so forth. Otherwise if the number of persons are not equal to the number of spatial terms and temporal terms, it assigns all persons to each location visit. The important thing is that a location visit can only have one spatial term.

In events where multiple expression have been found, there is high risk for errors and misplacements of information. The reference id for each sentence is also captured to be used in the data cleaning and organisation stage, which is aimed at confirming the extracted information against the data sources. At this stage, all the information are written to PostgreSQL database. The following screenshot shows the table for the Chronology of the Namibia history data.

	person text	location text	date text	temporalrelation text	spatialrelation text	sentenceid integer
97	Tjetjo	between Otjiku...	03.04.1904		between Otjikuara	3201
98	the Germans	between Otjiku...	03.04.1904		between Otjikuara	3201
99	Samuel Maharero	Okatumba	10.04.1904			3401
100	Samuel Maharero	Oviumbo	10.04.1904			3401
101	the Germans,Leutwein,Ovaherero	Oviumbo	13.04.1904			3501
102	the Germans,Leutwein,Ovaherero	Otjosazu	13.04.1904			3501
103	Von Glasenappâ	Otjihangwe	24.04.1904			3502
104	Von Glasenappâ	Otjihaenena	24.04.1904			3502
105	Ovaherero	Waterberg	19.04.1904			3601
106	the Germans	Engarawau	19.04.1904			3602
107	Ovaherero	Okangundi	28.04.1904			3701
108	Arthur Koppel	Warmquelle	20.05.1904		near Zesfontein	3901
109	Kutako	Tsumeb	06.08.1904			4803
110	Herero	Waterberg	10.08.1904			5002
111	Von Estorff	Okomiparum	10.08.1904			5003

Figure 15. Location events in Postgres table

At this point all the location visit events are written to the database table for geocoding and cleaning. The cleaning of the data involves date conversions such as 13 July to 13.0.1904, date type conversion and in cases where multiple troops are retrieved in a location visit, cross reference check is done using the sentence id.

3.5.1 Geocoding

Geocoding is a process of assigning geographic coordinates to location names be able to visualize them on a map. The location visit events are geocoded in Postgres by joining a location table which already contained geographic coordinates of all place names in the documents.

CHAPTER 4: HISTORICAL SPATIO-TEMPORAL DATA IN ARCGIS

4.1 INTRODUCTION

GIS applications are known for their capabilities in handling the spatial aspect of geographic data. However, a GIS with capabilities to process, manage and analyse spatio-temporal data has greater benefit in handling historical information because temporal and space-time attributes of spatial phenomena are quite difficult to represent and examine with current GIS. Particularly for ArcGIS, there has been tremendous improvements in the development of functions and tools to handle temporal information. In this section, we examine and evaluate the capabilities of the available tools for handling spatio-temporal data in regards to modelling, analysing and visualization considering the nature and structure of the historical information under at hand.

The next section describes the historical data in depth with regards to the temporal patterns and spatio-temporal data types.

4.2 SPATIO-TEMPORAL DATA

4.2.1 Spatio-temporal data type

Defining what spatio-temporal data type the data is depends on the overall goal on how this information is to be presented. Indeed, the same information arranged in different ways can serve different purposes. For example, information on troops at different locations in time, can be used for two different goals: 1. to model the trajectories of the troops as they move from location to location, 2. to model the locations visited in time. This translates the spatio-temporal information into two different spatio-temporal data types namely moving objects and location events in time. For this research the following spatio-temporal data types are considered:

- Moving points
- Point events in time

Moving point

To model spatio-temporal changes there is a need to keep track of the spatial locations. A moving object type represents objects whose location continuously changes in time and whose geographic extent is not relevant. For instance, the troop's movement data in this research are considered as moving object type, because the geographic extent of the location is not considered. In this case, a point feature represents a location as a function of time and can be mapped to a curve in the three dimensional space (x, y, time). The trajectories of the troops are then produced from the projection of the moving points on the plane by connecting two location and different time steps.

Considering the fact that time and spatial dimensions can be perceived to be continuous, the moving point type should allow identifying location and time at a random point inserted between two given time steps, making it easy to model trajectories for moving entities. This is substantially more efficient in human movement analysis, animal tracking and studies that deals with moving objects. Figure 16 illustrates the moving point theory.

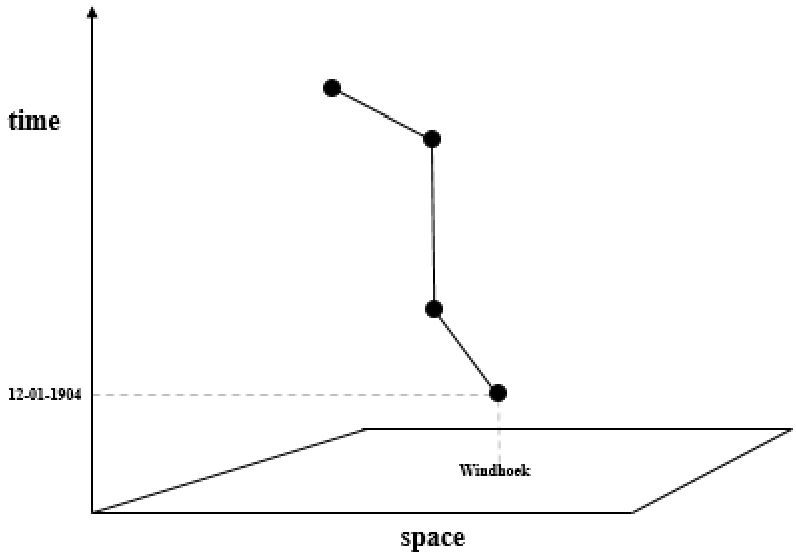


Figure 16. A moving point in space and time

As mentioned above, the overall goal behind the analysis has a great influence on the spatio-temporal data type. Moving regions is another spatio-temporal type that allows to model and track movement of an object with great emphasis on the geographic extent. This is more relevant for studies such as fire mapping where the extent of the geographic location needs to be addressed. The next section discusses the point events in time as another spatio-temporal data type for this research.

Point events in time

Spatio-temporal data can refer to any event information with time and location attributes. Historical information provides the opportunity to deal with discrete events that happened at different locations in a dynamic temporal manner. Events whose duration is not relevant and which might not repeat or happen in a continuous manner but are rather discrete at any point in time and are regarded as point events in time in this research. This type of events may be related or completely unrelated events. This refers in our case to battle events that happened through the year 1904 at different locations. The location information for all the troops together are also modelled as point events in time. Modelling these events as points in space and time in this research has a benefit of movement hotspots analysis by identifying spatial clusters which helps answer question such as “At which location where there more movements?” hence providing an overview of the concentrated areas. These events are represented by point features in space and time.

4.2.2 Temporal GIS patterns

Time and space are ubiquitous aspects of reality (reference). In order to model spatio-temporal data there is a need to keep track of the spatial location at a point in time, for a certain observation. The location and time of an observation can be fixed or varies in time. The spatio-temporal data is defined by the temporal pattern of the observation or to say the time intensity. Temporal GIS patterns defines the temporal behaviour of changing spatial objects which are either dynamic or continuous. Dynamic spatio-temporal objects changes continuously and static objects changes according to sudden events. Understanding the temporal GIS patterns of the data helps in the initial stages of spatio-temporal management in a GIS. The historical spatio-temporal information used in this research exhibits the following temporal GIS patterns:

Dynamic temporal pattern – an object is considered dynamic when its geometry or attribute changes over time. The duration of change is defined by the time fields in the data. Dynamic objects may be discrete or continuous.

Continuous dynamic information - usually applies when time is continuous and the object state changes continuously with time. For instance, some troop's information describes their movement on a weekly basis, where location changes at each time change in a short period of time and duration may be relevant during the analysis.

Discrete dynamic information – refers to objects whose state, location and time are discrete in nature. Changes in any of these aspects only occur at clock time when there has been any change. This refers to the battle events, whose locations and time are discrete and are not continuous. The battle information defines the location and time of individual battle events throughout the country, where by some location may have multiple battles at different dates. With discrete dynamic information in this regard, duration of change is not relevant.

Most events in real life are dynamic in nature (reference), and when modelled in GIS the following dynamic aspects of the spatio information is considered: (Spatio-temporal modelling of dynamic phenomena in GIS, pg 3):

- Change of feature's geometry over time – such as city expansion
- Change of feature's position over time – such as troop movement
- Change of feature attributes over time – such as temperature
- Combination of the above changes over time

Static temporal pattern – Although most phenomena in nature are perceived to be dynamic, there are static phenomena that describe objects that may be fixed and may not change over a short period of time. Such as roads and building just to mention a few may change only over a long period of time when there has been an upgrade. Static temporal pattern does not apply to the data in this research.

4.2.3 Requirement for modelling, analysing and visualizing historical data

Based on the historical information described in chapter 3, the following points are formulated as requirements to handle the historical spatio-temporal information in regards to the case study.

- To represent battles and location events in space and existence in time
- To capture the change in position in space over time for moving troops
- To represent change in position in a progressive way
- To represent the spatial relationships of troops in time
- To query semantics attributes of spatial objects in time - attributes that describes the trajectories such as the amount of time spent on a trip to the next location.

4.3 COMPONENTS OF A TEMPORAL GIS

Temporal GIS applications have given new ways to manage, analyse and visualize geography in time. A GIS application is made up different system components whose functionalities are connected together for effective spatial data handling. Handling time in GIS requires enhancement of conventional GIS application to accommodate the temporal dimension. This includes extension of spatial databases, development of tools to support temporal data analysis and visualization. The main temporal GIS components are *temporal databases*, *temporal analysis* and *temporal visualization*. The relationship between these components dictates the effectiveness in data handling, analysing and presentation.

The following figure illustrates the relationship between the temporal GIS components and their respective functions.

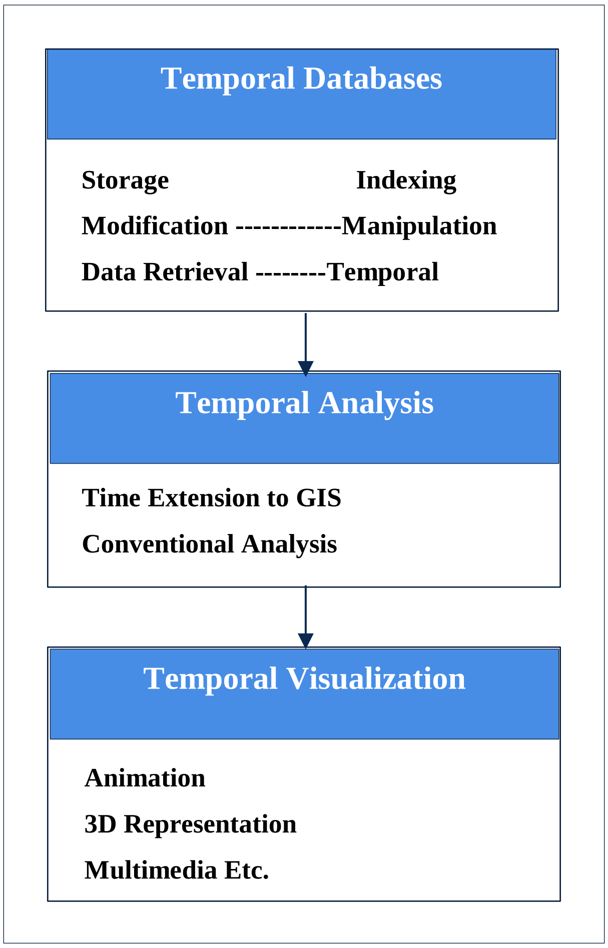


Figure 17. Relationship between Temporal GIS components

4.4 MODELLING AND PRESENTING TIME IN ARCGIS

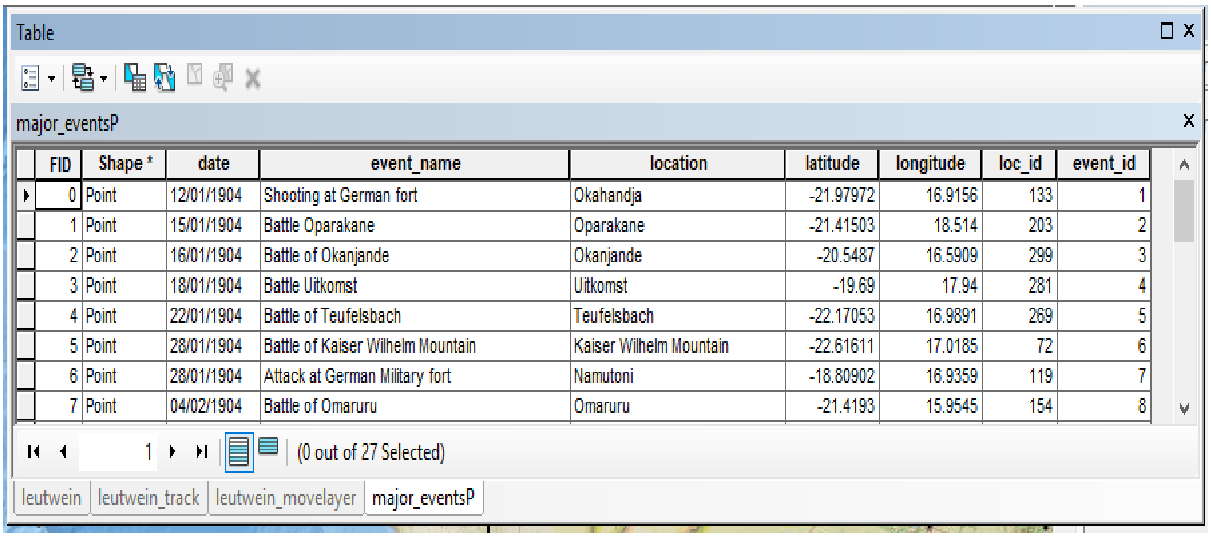
Temporal information in ArcGIS can be modelled as:

- Points in time
- Durations of time
- Transactional time

ArcGIS provides different ways to support time in spatial data. Time information can be stored as an attribute for feature classes, raster layers and tables or can be stored internally in data structured such as Network Common Data Form (NetCDF) and tracking layers.

4.4.1 Modelling historical events as point in time

Considering the scale of information used in this research, feature layers are used to represent the spatial dimension of the information. The battle events contain time information when it had happened and these are discrete events that happened in no particular order. What is required is the representation of event objects at existence time in space. This entails displaying a point on location at a time of existence, and each battle is stored as a separate feature (row) in the feature layer, with the time information as an attribute.



FID	Shape *	date	event_name	location	latitude	longitude	loc_id	event_id
0	Point	12/01/1904	Shooting at German fort	Okahandja	-21.97972	16.9156	133	1
1	Point	15/01/1904	Battle Oparakane	Oparakane	-21.41503	18.514	203	2
2	Point	16/01/1904	Battle of Okanjande	Okanjande	-20.5487	16.5909	299	3
3	Point	18/01/1904	Battle Uitkomst	Uitkomst	-19.69	17.94	281	4
4	Point	22/01/1904	Battle of Teufelsbach	Teufelsbach	-22.17053	16.9891	269	5
5	Point	28/01/1904	Battle of Kaiser Wilhelm Mountain	Kaiser Wilhelm Mountain	-22.61611	17.0185	72	6
6	Point	28/01/1904	Attack at German Military fort	Namutoni	-18.80902	16.9359	119	7
7	Point	04/02/1904	Battle of Omaruru	Omaruru	-21.4193	15.9545	154	8

Figure 18. Time as attribute for point in time objects

Figure 18 represents the attribute table for battle events, with time information stored in an attribute field. Each row represents a different battle event. Time information can be used in simple queries using the select by attribute SQL functionality.

4.4.2 Modelling troop trajectory as duration in time

Time is the indefinite continuous progression of existence (Wikipedia, 2017). Some events in life happen at a duration of time, where it is denoted with a start time t_1 and an end time t_2. Particularly in GIS, modelling progression of events with the start date and end date makes it easy for representation purposes. In this case the activities of a moving object cannot overlap temporally. The trajectories of the different troops are modelled considering the date a troop was present at a specific location represented as a point. Each point at a location represents one location stop at a time during the journey. The trajectory is made up of a successive set of location stops at irregular time intervals. For the purpose of visualizing movement, the gaps in time between location stops are misinterpreted as no existence. This introduces temporal gaps where the object is not displayed and only appears on the next time step. For example when there are 10 days between two successive location stops, ArcGIS cannot attain the state of an object before the next time step unless an object is stored with a start date and end date which is the date of the next location stop. This introduces a state duration which makes it easy to visualize the temporal information that changes at irregular time. To demonstrate the time gaps in the data between two location stops, Figure 19 summarizes in a nutshell.

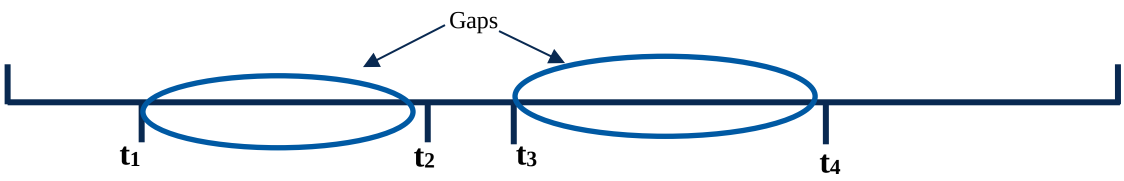


Figure 19. Temporal gaps at irregular spaced time steps

As mentioned earlier it is important to keep track of the position of an object in motion for the purpose of modelling movement. This helps in visualization. To solve the issue of visual gaps issue each location stop object is assigned an end date using the Calculate end date tool in ArcGIS. The end date of an object is the start date of the next object on the temporal scale. With this, each object is associated with a validity time and the state of the object remains the same or in regards to moving object, the object remains at the previous location for the duration time till the next location time. Figure 20 shows the calculated end time for each location stop and how it is stored in the attribute table of a feature in ArcGIS.

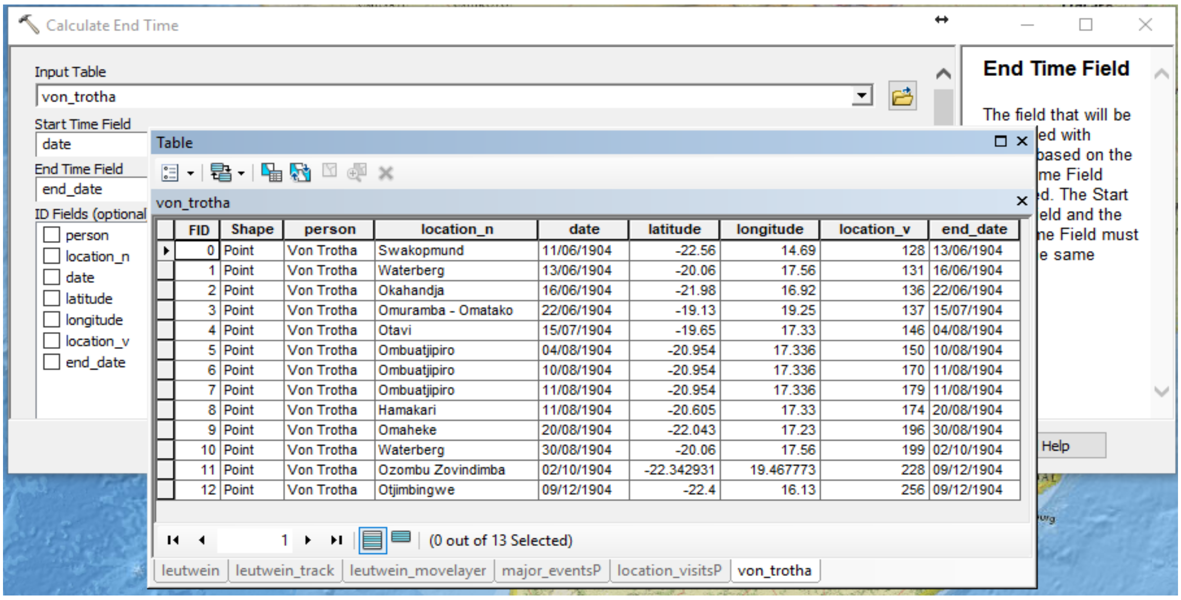


Figure 20: Calculated end time

4.4.3 Date field format in ArcGIS

Date data type – Time is dynamic and has so many temporal units such as a year, which contains 24 months and each month can have up to 31 days. When working with time values in ArcGIS it is recommended to use a homogenous sampling unit such as date, year or hours etc. In this case time can be stored as Integer, String, and Double or Date format as recommended. Date field format is a special data type specific to time, which makes it easier to access, query time and use geoprocessing tools in ArcGIS that support time information.

Date formats – Date formats also play a role in the representation of time in a table. ArcGIS supports numerous date formats such as “YYYY/MM/DD hh:mm:ss” for Year, Month, date, hours, minutes and seconds. For this research, the used date format is “YYYY/MM/DD” which has been automatically converted during migration from PostgreSQL to ArcGIS from “DD/MM/YYYY” as stored in Postgres. However, ArcGIS provides tool to convert date from one date format to another when needed.

4.5 ANALYSING HISTORICAL SPATIO-TEMPORAL DATA IN ARCGIS

Temporal analysis allows us to answer question considering time in relation to space. Temporal analysis give us the benefit of spatio-temporal data in general. However, the choice of tools and functions to use for temporal analysis in GIS depends largely on the scale of the data, the underlying objectives and motivation. The motivation of carrying out temporal analysis in this section is to find out the analysis methods that supports historical movement data in small scale with poor temporal density. With an interest to explore the geographic content of the Herero war of resistance by studying the time-geography setup of the different events by answering spatio-temporal questions pertaining to “where”, “when” and “what”. Moreover, the main objective is to understand the concept of spatio-temporal data in GIS applications and how to support the handling of such data.

Spatio-temporal data describes the “what”, “who” and “when” of a historic event. An example, finding out who was present at a specific battle events in a space and time or analysis related to the state of an object a specific point in time. To answer spatio-temporal questions of this nature, we consider the scale of the data, the temporal patterns and the analysis questions that need to be addressed.

In regards to the moving object data that represents the set of locations on each troop’s journey, being able to answer questions related to their stay at certain locations is of great importance at this stage. Capabilities to also analyse the movement patterns for a collective of troops at a point in time would also be of great significance in this section too. We would also like to analyse the spatial connection of movement events in time along with the battle events. The question to be addressed in this section are: Can we deduce any spatio-temporal reasoning in regards to the time and location of the historical events in our data? Are there any spatio-temporal clusters in movement data? Is there any benefit to spatio-temporal modelling and analysis of these information? Is our data sufficient and suitable for spatio-temporal analysis? To unveil the patterns and answers these questions, the following questions are used to assess the capability and functionality of ArcGIS framework to analyse spatio-temporal data in regards to our historical information.

1. At which areas were there more events?
2. During which time period did the people move more?
3. During the battle events, who was close in space and location?

4. Any relationship between the battle events and movement of different groups?

These questions determine the analysis methods and tools to be used. Are there any functions and tools to answer these questions?

4.5.1 Spatial – Temporal Clustering Analysis

Spatial clustering in GIS plays a big role in quantifying geographic variation patterns. It is often used in other fields such as epidemiology, crime analysis and population studies (Jacquez, 2008). The fundamental principle is the same in all these fields, which is to map the excess or count of events or object values at a location. While spatio-temporal clustering is a process of grouping objects based on their spatial and temporal similarity. The analysis of spatio-temporal data in context requires both temporal and spatial correlations to group events that are close in space and time. Therefore, in this section by using the spatio-temporal clustering method we intend to derive and get an insight on the locations which had more events happening compared to others to answer the question “At which areas were there more events?”

For this we considered the following tools in ArcGIS:

- Spatial Statistics – Generate spatial weight matrix tool
- Grouping Analysis tool
- Isarithmic method

4.5.1.1 Generate spatial weight matrix

If we are able to define spatial neighbours in time, traditional spatial analysis becomes space-time analysis. ArcGIS spatial statistics tools allow us to model spatial relationships between objects in time. The Generate Weight Matrix tool is used to generate, store and share the formation of the relationship among the objects in a data structure called .swm. The conceptualization method specifies how the spatial relationship among the objects is carried out. Space-time window conceptualization method is used with a defined temporal interval and a cut off distance, to group features close in time and location. Using the location visit events, we try to generate spatial-temporal neighbours looking into a 4 weeks’ time frame and within 50 km considering the landscape constrains and distances between Namibian localities.

The generate weight matrix tool returns a .swm matrix file which is converted to table by using the convert matrix file to table in ArcGIS. This table is used in the Grouping analysis tool under spatial statistics tools to group the objects according to the specified neighbours in the swm file. We are interested in finding the spatial clusters of location events within 50km distance and which happened within a month time. However, our data proved to be insufficient in terms of spatial and temporal correlations. Most events do not have neighbours in time nor location at the specified parameters. The following Figure 22 and Figure 22 show how to generate weight matrix, and the results of the tool.

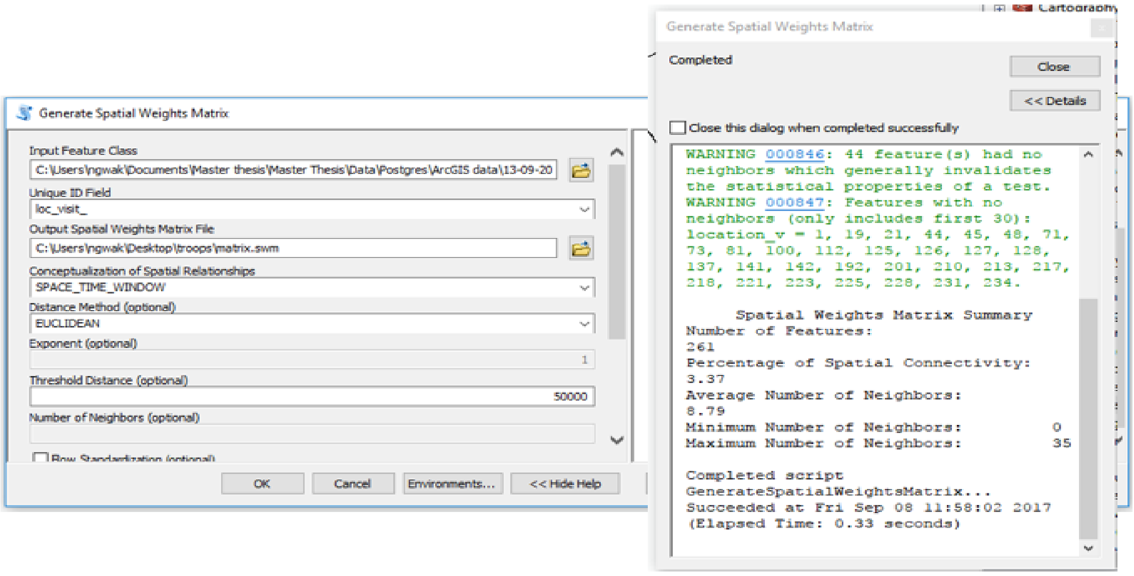


Figure 21. Generating Spatial Weight Matrix

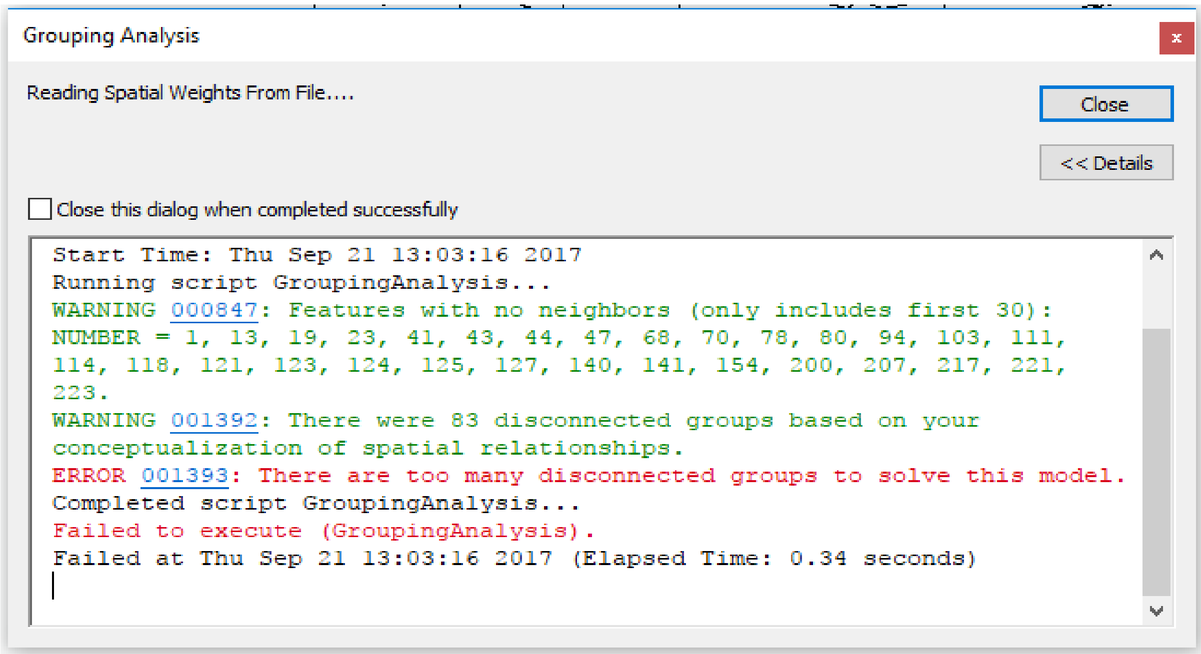


Figure 22. Grouping Analysis Error

Figure 22 shows the results of the grouping analysis tool as attempted. There are many disconnected events at 100km, 200km and 300km distance as well and time and therefore it was not able to group features as desired.

4.5.1.2 Isarithmic Map

The idea is to do hotspot analysis method which uses statistical analysis to identify areas of high occurrences and low event occurrences. Hotspot analysis defines the hotspot areas that are statistically significant. Unfortunately, our location event data is not sufficient for the hotspot analysis function in ArcGIS. However, we can use the Heat map function to create the isarithmic map visualizing the locations with significant amount of events. Since each row represents an event at a location in time, events happening at the same location will overlay each other. The heat map method takes the point data and create an interpolated surface which shows the density of occurrences throughout the geographic study area. The isarithmic map is created in ArcGIS Pro and visualized in time using the time slider.

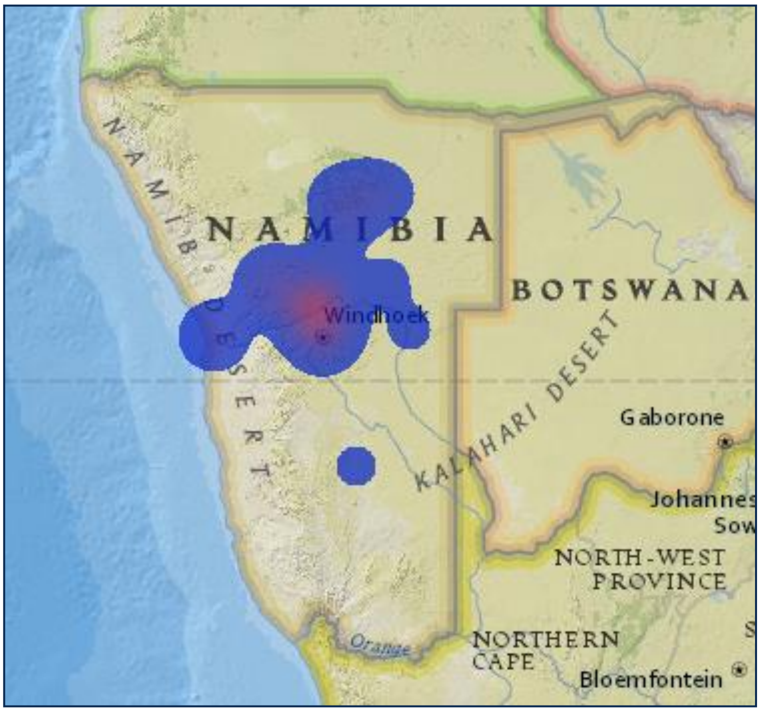


Figure 23. Heat map in ArcGIS Pro

The interpretation of the isarithmic map results is said to be subjective to the reader in comparison to the hotspot maps (Dempsey, 2014). In this example, the red areas show the concentrated areas and the size shows the extent of the area covered by all location event points.

In the same manner, we would like to answer the question of which temporal period constitutes more movement events. And because the historical data for this research is limited in time and space, we aggregated the location visit events into monthly epochs to have a substantial amount of information to analyse. The heat map method similar as the section above, is used to display and help detect aggregated movement patterns on a monthly basis. The figure below shows the static maps for each month from January to December that depicts the movement patterns. This gives an indication on which month had more movements according to the publications and the spatial locations. Additionally, animations are used to stimulate change detection and for ease display.

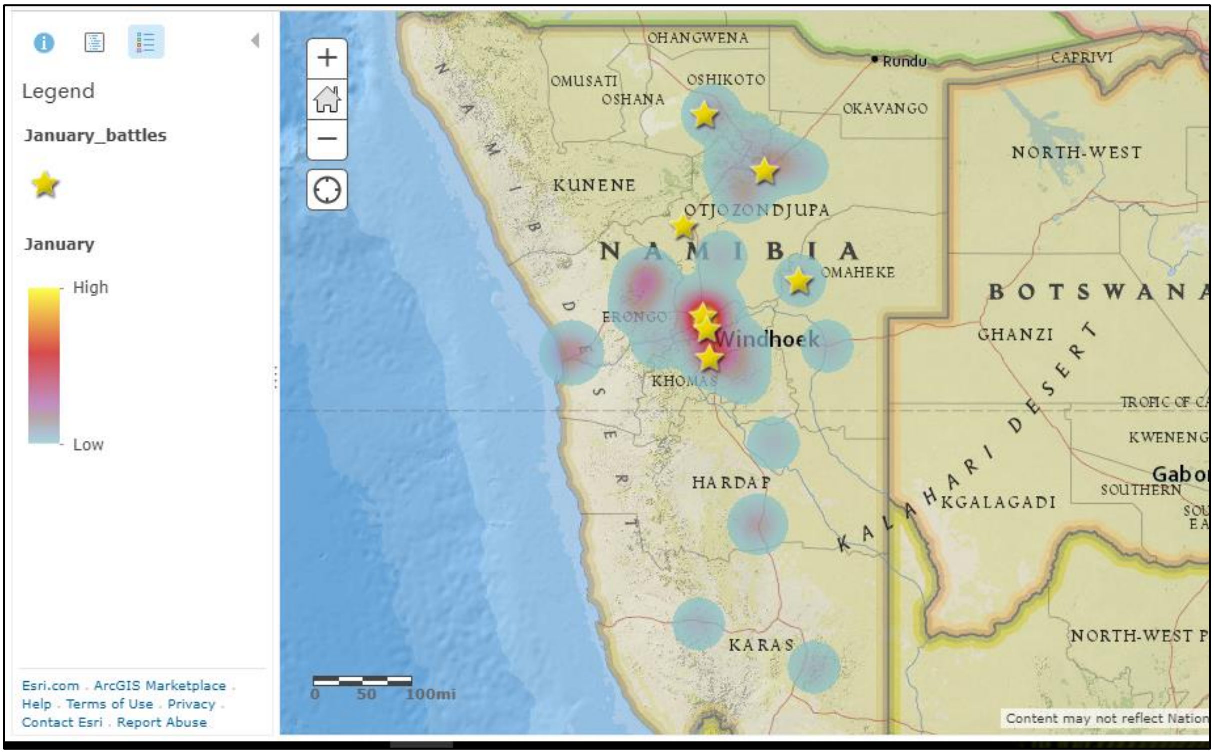


Figure 24. Monthly event analysis

The events illustrated in Figure 24 includes the battle events in yellow stars and the isarithmic visualization for the month of January. The colour intensity for the represents the density of movement events at a location. Visualized together with the battle events for the same month, we can conclude that there has been more movement round about the central areas. These movements could be triggered by the battle events. Clearly, on the locations with less movement events (shown by the lighter blue colour) there hasn’t been any battle according to the data. This method allows us to draw relationships between the movement data and the battle events. However, this has no temporal benefit as the feature is regarded as time layer in space and is only suitable to visualize one time layer on a map. This is the reason why time animation is useful in this sense.

4.5.1.3 Trajectory Analysis and visualization

Moving troops trajectories are modelled as point data, extracted from the location visit events. Analysing point trajectories for historical events can be a challenge because some trajectories are short, very distant location and discrete times. However, there should be a way in GIS to model and represent trajectories regardless of the characteristics. A trajectory aims to show the journey an individual took at some points in time.

ArcGIS provides the option of creating trajectory line paths from moving points by connecting points in a temporal order. Apart from representation, the interest in modelling trajectories include analysing the movements in regards to time. For example, querying the moving object in motion, such as *"On a specific time, where was the object on the journey?"* The idea is also to be query moving object in relation others in time and space like, *"Where did the two object intersect in time and space?"*. However, the possibilities to carry out such analysis is rather not possible in current ArcGIS.

On the other hand, the display of trajectories paths as straight lines gives an indication of the direction of the journey. And by using time animations, we are able to view the movement in a temporal order. However, the representation raises questions regarding the moving and representation of time on the path. Some paths took more days than others, but the whole path is displayed in one go. To improvise and introduce a progression motion like approach, we manually created the paths using the following steps:

1. Create line paths connecting points in temporal order
2. Calculate time interval between two successive locations in days
3. Divide the each path line into equal segments
4. Divide interval days by the line segments
5. Assign new start and end time values to each line segment

The created trajectory paths are displayed using time slider to visualize movement in a motion approach.

We assume that the time difference between two locations could mean that the troop remained in the previous location for that amount time or the time it took for them to reach the next location. Duration of travel is of much significance in human movement studies and most especially in studies such as traffic movement and animal tracking. The idea is to be able to associate the trajectory paths with their attributes. For instance, visualizing the trajectory paths with longer duration of time and visualizing the paths according to the time spent on the trip. A similar approach can also be applied to the location points, to get an overview on where a troop had stayed longer.

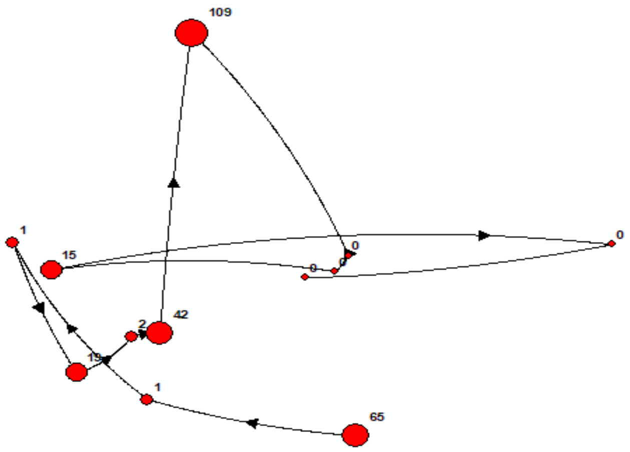


Figure 25. Trajectory attribute analysis

Figure 25 shows a trajectory and the duration of stay at each location illustrated by the size of the point. While this works for straight trajectories, most troops have actually been moving around the same locations for a couple of times. Analysing trajectory features in this sense may not be suitable for such trajectories as there are spatial overlays.

Trajectory analysis may be more interesting when analysing multiple trajectories together. In this case we are able to analyse the interactions when different moving objects meet in space and time (Shawet al., 2008). Like the STT data type from Frihida et al. (2009) has operators and functions to query on the semantics of STT type as well as spatial topological and temporal operators that would be useful in trajectory analysis. For example, it would be interesting to analyse activities and states of objects before or after a certain trip.

4.5.1.4 Space - time cube analysis

The time-geographic framework has been perceived as useful analytical framework to analysing and representing individual activities in a spatio-temporal perspective (Shaw et al., 2008). The 3D coordinate system and functionality in space-time cube method makes it easy to represent time in the representation and not have to impose it on the dataset with animations for instance.

In this section, we used space-time cube analysis tool in ArcGIS to analyse the spatial and temporal patterns of the historical events. Location events datasets contains the aggregated troop's movement information, whereby a row represents one location visit event. For instance, all troops present at the same location and same date belongs to one location event which is represented as a point in time. With the space-time cube we would like to analyse the collective movement patterns. Such as which location had more location events and which temporal term had more location events, similar to the isarithmic map. The space-time tool aggregates the points into 3D space – time bin structure. Each bin represents a number of points that fall within that distance in a specified temporal range.

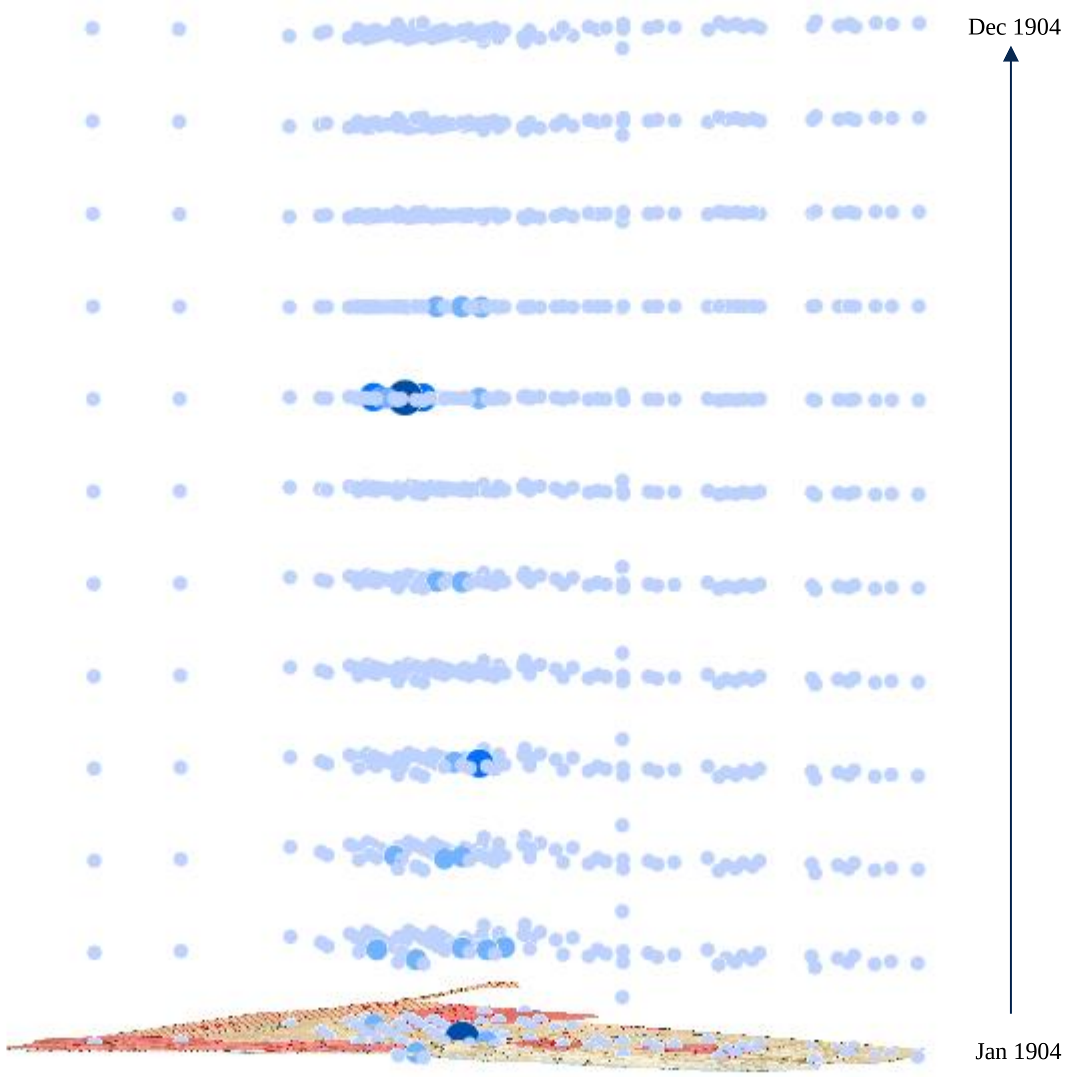


Figure 26. Space time information cube for location events

Figure 26 shows the 3D space-time cube for location events. Each layer represents location visit events per month aggregated at 5okm distance. The size and colour of the point illustrates the number of location visits at a location. The bigger and darker the point, the more the location visits. To clearly analyse the space cube analysis, it should be visualized in an interactive environment.

4.6 VISUALIZATION OF SPATIO-TEMPORAL DATA – CARTOGRAPHY POINT OF VIEW

Cartographers have been struggling with the concept of time for as long as maps have been produced (Kraak, 2003). The main issues to consider are the map production time and modelling the changes in real world time. Cartography aims to present design rules and methods that helps communicate changes in time in an effective way (Kraak, 2003). For example, illustrating changes by using cartographic graphic variables. An ideal visualization method for historical data should facilitate the story telling, presentation of the historical event as a process and considering the chronological order, spatial position and attributive information.

The objective of visualizing historical data in this research is to analyse and understand the natural order of the events as they happened. It is therefore important to make use of visualization techniques that displays the chronology of events and interactivity along with attributive information. For instance difference in time and number of troops at a location that facilitate the understanding and analysis of the events as a historical process. The idea of having flexible tools that supports time aggregations and creating map animations of the time aggregates is another factor considered in this research.

4.6.1 Visualization

As mentioned in section 4.5.1.4, space-time cube visualization provides the benefits of both worlds. By combining the 2D representation of space with time into 3D. The representation of spatio-temporal data in space-time cube can be either static or animated, however animation is not required to represent time. For the purpose of visualizing trajectory paths, 3D visualization best fits the objectives of visualizing movement in space with no overlaps. Space time cube visualization concepts are used in this research. See section 4.5.1.4 for more explanations.

4.6.2 Animation maps

In addition to all analysis methods in the previous section, this research embraces the time animation methods in order to present and visualize movements at different times. This applies to trajectories, 3D space cubes and heat map visualization in section 4.5. The challenge of showing time is no longer difficult with the support of software developments. ArcGIS 10 provides the benefit of creating animation maps by setting time properties on temporal features through animation manager or time properties. Additionally, time slider function also allows users to configure the time display by controlling the speed of display and setting the time steps. Users can also stop the animations at a particular time on the slider to probably get an insight into the visualization. The advantage of time slider is the interactivity measure it provides to users in controlling time change displays.

Figure 27 shows the time settings and time slider concept in ArcGIS:

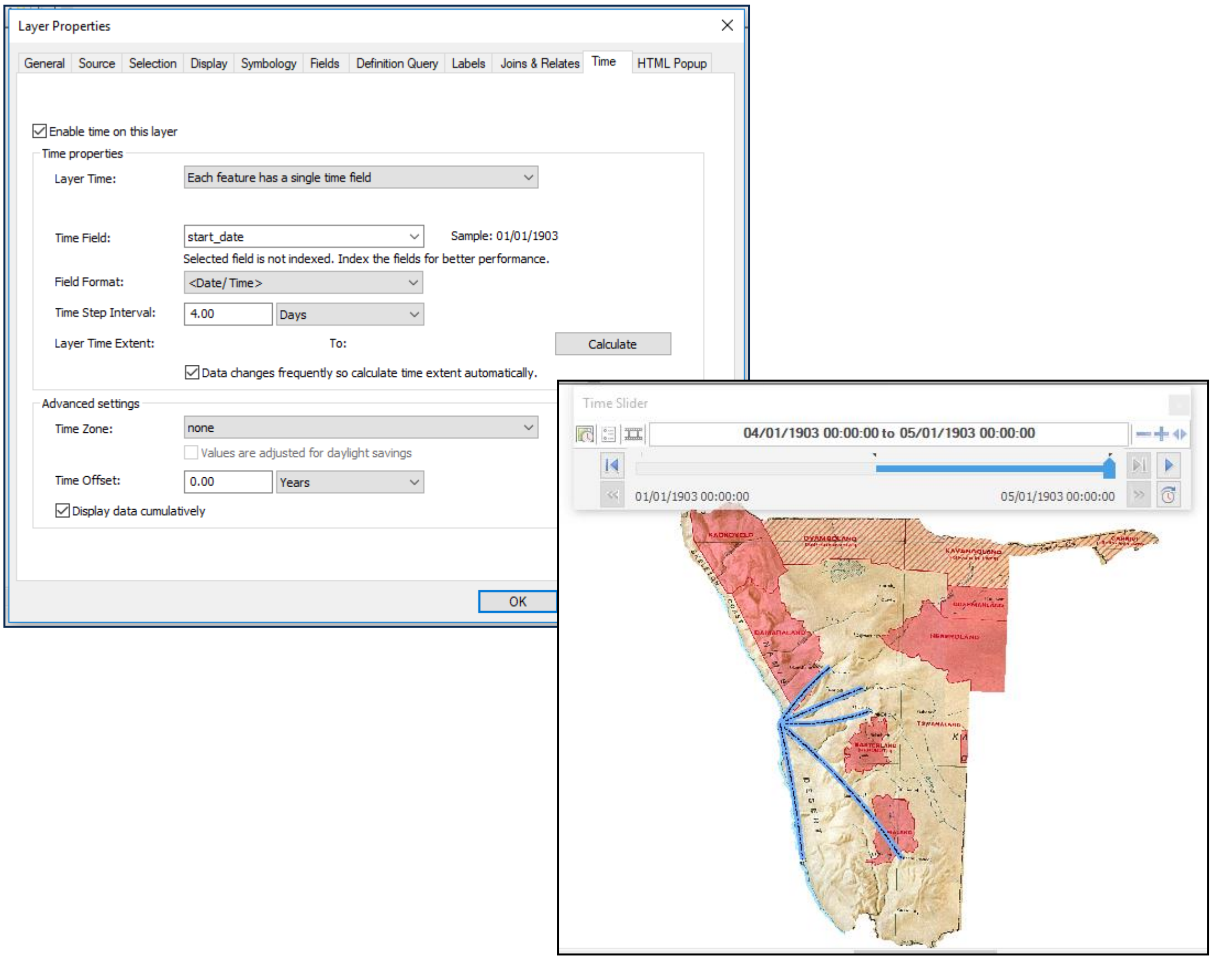


Figure 27. Time slider in ArcGIS

A time aware map is created to illustrate the movement of the German troops before and during the battle of Hamakari on the 11th August 1904. Time animations are used to show the progressive movement of the respective troops as they advance towards the Herero camps at Waterberg Plateau. This event is chosen for the trajectory movement visualization because of the short temporal extent, which is easier to configure compared to representing discrete trajectories with different time extents. See Figure 30: Time Aware story map - Battle of Hamakari 1904.

However, time animations using the time slider do not solve all time change display issues yet. It is difficult to compare changes in for example two objects located a distance further from each other. It is also on this note that more cartographic animation variables are considered in order to convey time changes in spatio-temporal data visualization. The variables of animation in cartography can include graphical manipulations and sometimes sound effects. To depict change in location events from time to time, size, colour and position variables are used to emphasize the more concentrated areas and change in position. Time animations are added to accentuate change in time.

- Size – The size of symbols in point data used to represents the amount of troops at a certain location at a point in time. The size is made proportionally larger or small to show the amount of location visits in time at a location.
- Colour – Graduated colour symbology is used in isarithmic maps to represents the concentrated areas during the events from time to time.
- Position – Change in position of a point data, or line represents to movement.

4.6.3 Story Map visualization

A story map is a dynamic visualization technique that takes a user centric perspective to represent map activities in a more narrative manner. Representing temporal historical information on a static map can be challenging because each point in time would require a map of its own. Ordering the static maps together dynamically to represent different events, allows us to link the events together and introduce a temporal perspective at viewing. The benefit of using a story map in this research is the ability to link maps, incorporate images and the flexible user control functionalities that provide the opportunity to see time events at the pace of the viewer. Esri story map journal was used to create a series of maps dynamically linked together to represent the order of historical events during the Herero Uprising was 1904. With the benefit of time representations, the time aware map created in section 4.6.2 is dynamically linked into the story map to illustrate the movement events during the decisive battle of Hamakari. The story map is created and hosted on ArcGIS online.

The frame of the story map contains a side panel, main content window, navigation slider on the left. The side panel contains the texts explaining the contents in the main panel. We embedded source maps, photographs, and timelines in the scrolling narrative on the left side of the map or as a substitute for the web map in the main content window. This story map contain three main sections: the introductory section, the historical events which contain dynamic linked maps of the battle events and the time aware map for the battle of Hamakari on the 11th August 1904. The figure below shows the introductory page of the story map with image contents.

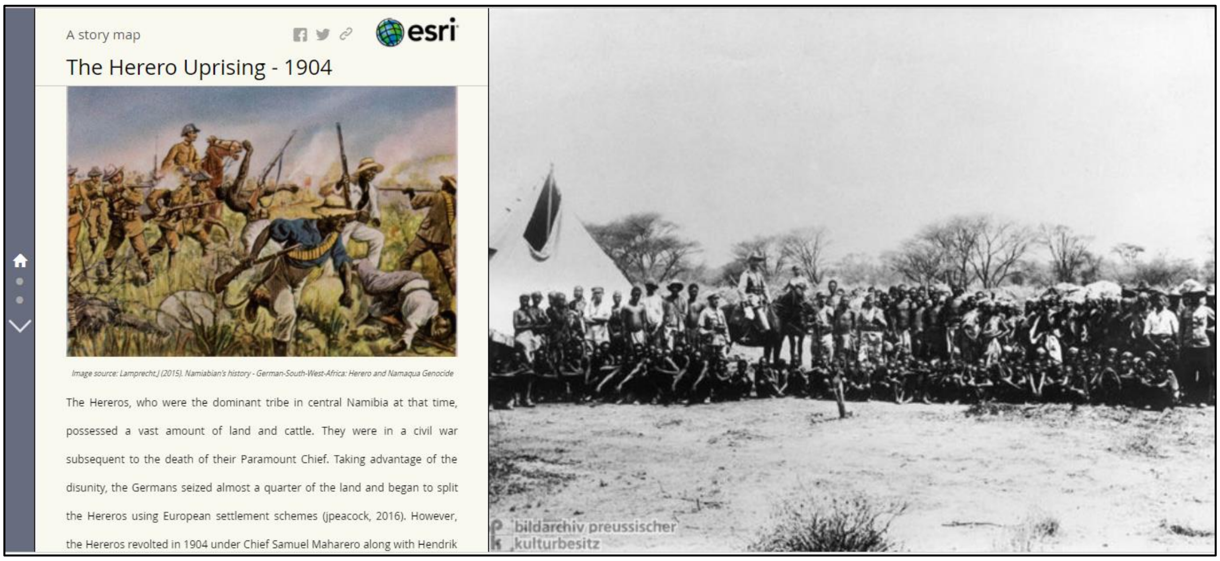


Figure 28: Screen capture of the "Herero Uprising" introductory page of the Story Map

The side panel provides enough room for text descriptions. The images on the side panel can also be enlarged at view when clicked. The second section of the story map contains the representation of the battle events in a chronological order and dynamically displayed when a user clicks on the battle list in the side panel, texts are synchronized with the map contents to display features of interests. Clicking on the list of the battles in the text description changes the map extent to display the location extend of the clicked battle in the map. An overview map is placed at the bottom left corner of the main content window to show the location at the country level. This is another way of visualizing discrete historical events in this research in a chronological order without using time animation.

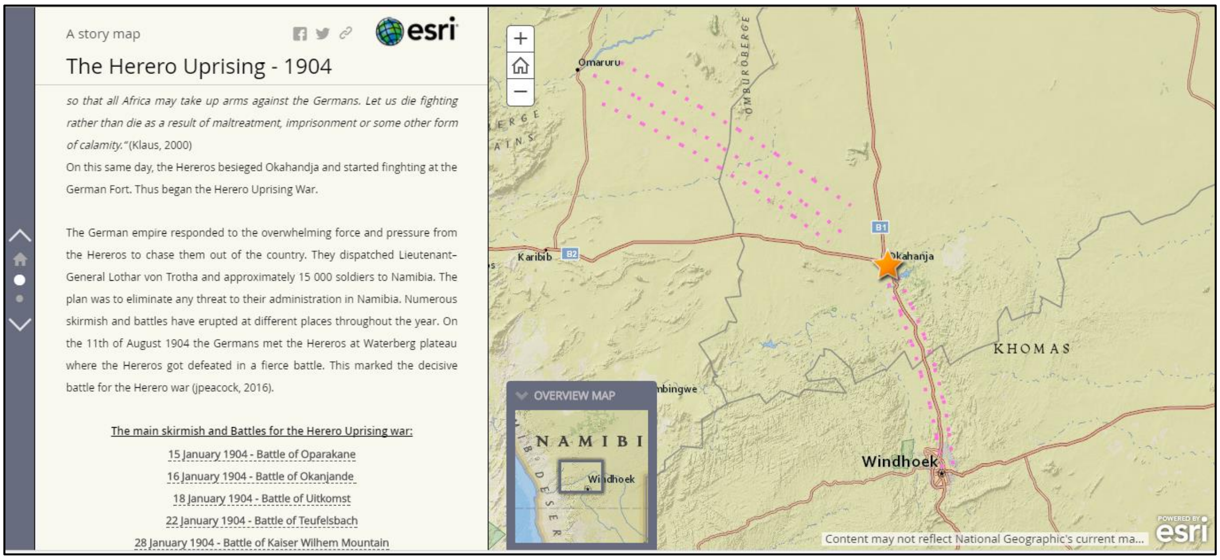


Figure 29: Dynamic story map - battle events

The time aware map shows the movement of the German troops before and during the battle of Hamakari on the 11 August 1904. A number of six troops had fixed positions on the fifth of August surrounding the Waterberg Plateau, where the Hereros had their camps. From the fifth, they advanced towards the Herero camp from different directions. On the 11th each troop stood at a position close to the Hereros and the battle began. The time map contains the time slider, which provides animation controls and allows users to pause and zoom to feature of interest at that point on the time line. Along with the time map, the texts on the left narrates the map situation for ease understanding and orientation. Figure 30: Time Aware story map - Battle of Hamakari 1904.

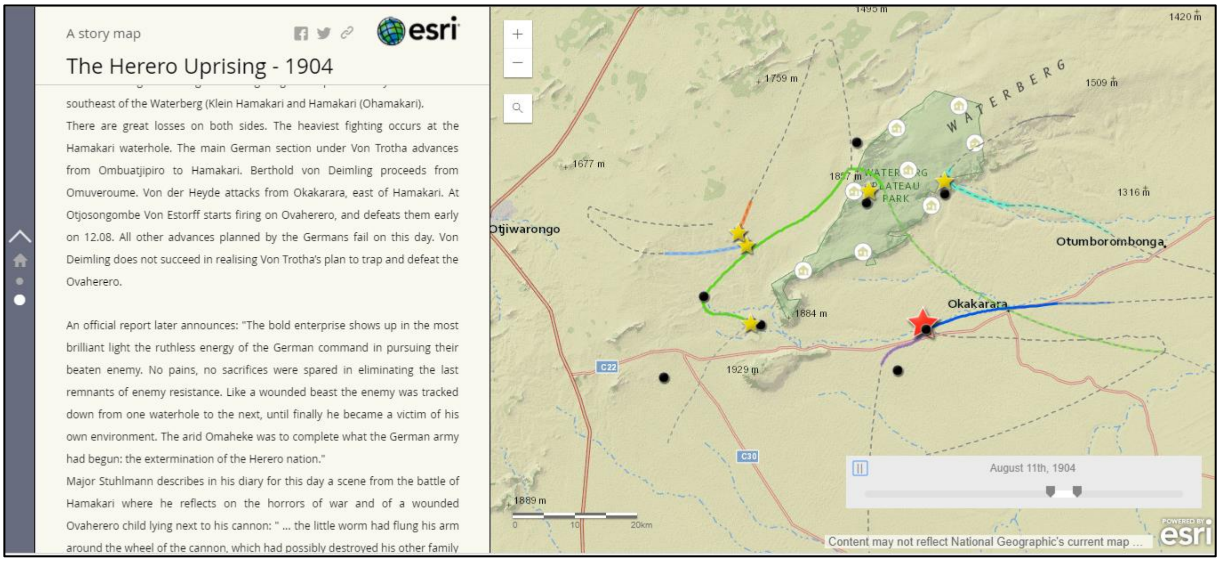


Figure 30: Time Aware story map - Battle of Hamakari 1904

CHAPTER 5: EVALUATION

This Master's thesis consists of two main research objectives, (i) to find suitable methods to recognize and extract historical spatio-temporal data from text documents and (ii) to model the extracted data and evaluate the effectiveness of the ArcGIS tools in handling historical temporal information. This chapter discusses the effectiveness of the used tools and methods as well present the shortcomings of historical spatio-temporal information.

5.1 EVALUATION

5.1.1 Methods for historical data extraction

The extraction and representation of spatio-temporal information relating to historical events constitutes the heart of this thesis. The extraction methods described in this thesis explains how domain specific information is captured and extracted from texts by using Text Processing methods, Gazetteer matching methods and JAPE grammar rule. GATE development application provided the framework to develop an application to automatically recognize and retrieve spatio-temporal information along with the attributive information as discussed in chapter 3. Not only did these methods provide spatio-temporal information, but also the ability to retrieve implicit spatio-temporal terms known as spatio-temporal relationships. This is all possible because of the flexibility of GATE application and availability of the NLP resources within GATE. One of the strongest point of this method is the ability to customize the Processing Resources such as the extension of gazetteers to include the information being extracted. Additionally, the functionality of JAPE grammar rule method allows maximum potential extraction for desired information. However, the effectiveness of JAPE grammar rule is fully depended on the definition of the JAPE rules and gazetteer lists of entities. If the annotation classes and rules are not well defined, the annotation process may not be successful. Unlike other methods, GATE framework provides the annotations in the order of the document which makes it even easier to extract and combine the references together.

For the purpose of this research, and as mentioned earlier, we were able to extract spatial terms, temporal terms, attributive information such as people's names and spatio-temporal relations separately. These references are combined together using an algorithm developed in python that uses the GATE document order to combine information belonging to one event in a chronological order. The combined information formed location events and trajectories for different troops.

5.1.2 Modelling time in ArcGIS

The character of abstract geospatial data types enables and constrains the kind of possible representations in GIS. This is not an exception when dealing with time data. While there are a number of data models designed specifically for spatio-temporal data representation, in this research we evaluate the effectiveness of representing time on an abstract data type.

As mentioned earlier, there are 3 spatio-temporal phenomena in this research. (i) Moving troops, (ii) battles events and (iii) location visit events. These phenomena provided the benefit of time information and therefore their time attributes is of great significance in this research. Using ArcGIS 10, we evaluate representation of time, the formats and way to retrieve such time information. The idea was to find out how much we can do with time information on an abstract data type.

The moving troops are point features that represents the location points a troop had been. Time in this sense is modelled as a point in time, where by time information is an attribute of the location point. Same applies to battle events and location visit events. Time stored as an attribute of a feature provides very limited functionalities such as simple time queries. In addition, we can still access time information for other geoprocessing tools such as space-time pattern mining tools and also setting time properties for visualization. Figure 31 shows how to access time information through simple time queries in ArcGIS Pro.

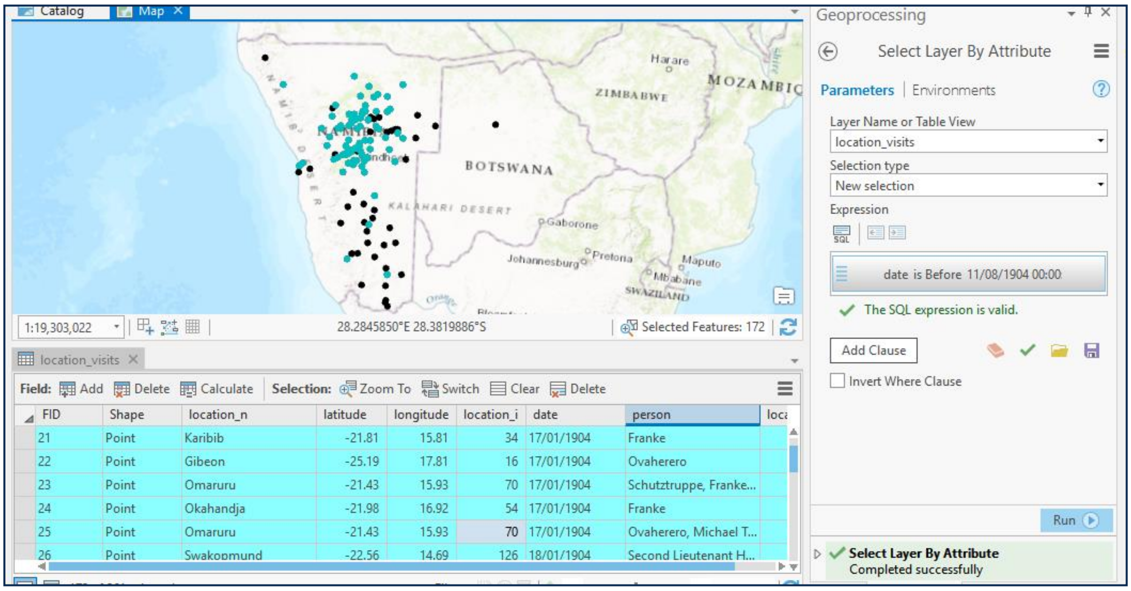


Figure 31: Simple time queries in ArcGIS Pro

Moreover, ArcGIS provides functionalities to store temporal data in separate tables. Where by the spatial information is stored in one table and values changes over time are stored in another table. You would use join operators to access this value changes. However, this is not covered in this research as it is more suitable for multivariate temporal data such as weather data. Additionally, the spatio-temporal phenomena for this research contains dynamic time information and very irregular for this kind of time modelling.

In regards to retrieval of time information, the abstract data type restricts the complex operations on time. Apart from calculations of end time, complex functions are lacking to support for instance topological relationships in time. This hinders the ability to carry more insightful analysis. However, this explains the reason for the growing research interests on spatio-temporal data models in the GIScience field.

5.1.3 Analysis and visualization tools

Spatio-temporal analysis aims to answer questions regarding the occurrence of events in space and time, as well as the relationships among them. With the aim of answering the analysis questions presented in chapter 4, different analysis tools were used. In this section, we discuss the functionality of these tools, the expectations and the results.

ArcGIS provide Space Time pattern mining toolbox that contains statistical tools for analysing data distributions and patterns in the context of both space and time. Space-time pattern mining are of great significance to this research to help answer the analysis questions. We evaluate the space-time cube tools in the next sections.

5.1.3.1 Space time cube tool

Space time cube analysis tool combines time and location in a natural way (Li, 2005) and time can be represented as continuous or discrete. The aim for using this tool, was to aggregate the location visit events into space and time 3D representation, to illustrate the amount of troops at a location at different times. As explained in section 5.1.3.1, we aggregated the location visit events at a distance of 50km per month to get an insight on how much movement happened in each 50km distance from January to December. The space time cube results provided the answers to the analysis questions. However, apart from time and spatial aggregations it does not provide semantic representations. The space-time cube represents the aggregated points as counts of occurrences but not additional information regarding these points. For example options to visualize other attributive information that describes the features. It would be of great representation to count all location events in the specified spatio-temporal distance and further determines who these location visits belong to. The idea is to distinguish the German group from the Hereros as modelled in the attribute field of the location visit events. This would be of great significance to give in sight on where the respective groups met in time.

In addition, from a geo-visualization point of view, dynamic visualization and interactivity play a bit role in 3D visualizations. At the moment the space time cube in ArcScene allows views from different angles and coupled with time settings we can create time animations too. However, having the functionalities to query the cube’s content is limited since the space time cube does not carry the semantics from the original feature layer.

All in all, the space-time cube analysis results proves to answer the analysis questions. In correspondence to the battle events, we can deduce conclusions regarding why there has been more movements at certain locations and during a particular period.

5.1.3.2 Grouping analysis

As mentioned earlier, defining spatial neighbours in time, traditional spatial analysis becomes space-time analysis. Another tool in ArcGIS is the Generate weight matrix tool used in this research that creates spatio-temporal neighbours. The conceptualization method used was space time window with a temporal interval of a month. The results of the generate weigh matrix tool is table used by the Grouping analysis tool and group the feature according to the spatial weights contained in the table. This function sounds similar to the space time cube concept apart from the fact that it does not have 3D functionality. It also require substantial amount of points to generate spatial weights. For the purpose of grouping location events that happened close in time, the data quality proved insufficient for this tool.

5.1.3.3 Trajectory representations

The representation of moving trajectories as point data is rather inefficient for historical data. This is especially because of the discrete locations spatio-temporal pattern found in the data. ArcGIS tracking analyst tool provides the option of displaying track lines from point data. While this function allows us to visualize the movement in a linear perspective, it is not very useful for the historical data at hand. Some location points could be long distance away from each other and also many days away from each other. From visualization point of view, displaying movement time as a straight line is unnatural. There is however no tools that customizes the representation of paths in a motion like approach. Assuming that the difference in time between two are successive locations means the day it took to get the next location, there should be a way to automatically estimate the time and locations at each point on the line. To do this manually can be cumbersome for longer and many trajectories. Having such functionalities would help in visualizing movement progression instead of a straight line at once.

On another note, trajectories can contain spatial overlay when objects follows the same path repeatedly. This is the case with most trajectories in this research. To solve the overlaps, there is only time animation options. The lacking functionality of 3D space-time path hinders the visualization of trajectory lines on a static maps.

5.1.3.4 Time animations

Time animations have become the breakthrough for temporal data visualization. It allows the display of features at the time of existence. Overall, this is a great way of dealing with temporal overlaps and also engaging viewers to see the time events and notice changes. However, the display of events with dynamic discrete time patterns is compromised due to the limited time setting options. The time step properties defines the events displayed at a time. In our case, the battle events happened only 26 times throughout the year. Some battles happened close together in time and some at discrete times. In this case visualizing the events only on the date of existence is not an option. You can either have a daily time step that represents each day of the temporal range even when there are not events, or calculate an average time step which displays all the events within one time steps. This works well for large amount of events, but in cases of visualizing battles events it is rather not good. These could be important historical events that should be represented on the exact date of occurrence.

5.2 UNCERTAINTIES IN HISTORICAL SPATIO-TEMPORAL DATA

A historical method is made up techniques and guidelines by which historians use primary sources and other evidence, such as archaeology to research and write histories in the form of events of the pasts (“Historical method,” 2017). Different historians uses different historical methods to narrate historical findings. The method to use for transformation of historical information in to GIS representations is an important as aspect to consider for the data quality. The saying, “Garbage in, garbage out” applies when data full of errors, inaccurate and imprecise is used during analysis. However, the nature of historical data itself comes with natural uncertainties that is can be hard to correct.

The power of GIS lies in its ability to integrate datasets from different sources, related to the same geographic area and able to perform analysis. Bringing a new dataset in GIS, the software also imports the uncertainties that comes with the datasets. The first action to taking care of the problem of uncertainties in being aware of it and understanding the limitations of the historical data being used. This section discusses the limitations and uncertainties in the historical data as well as the measures taken to improve the accuracy of the data.

5.2.1 Positional uncertainties

The historical information used in this research describes events that happened in the 1900s. The landscapes and other features such as roads and settlements have changed from then. The historical information used contained mainly three issues relating to geographic positions of the events. Whereby, (i) Settlement names that do not exist anymore, (ii) location indicators that only provides an approximate indication of the location and (iii) uncertain geographic positions.

5.2.1.1 Settlement names that do not exist

Due to developments and the war of liberation that occurred after the German war in South West Africa now known as Namibia, people moved across the country to settle somewhere else. Some settlements were abandoned and in the process eventually ceased to exist. These places are only found in the historical publications and the map sketches. The map sketches gives an indication of the approximate locations. To handle this issues, the map sketches are digitised and georeferenced in order to obtain the uncertain locations. Do to that, the following measure are taken:

a. Identifying uncertain location names from the data
b. Identify the uncertain location points on the sketch map
c. Georeference the sketch maps
d. Identify common control points from the reference layer and the sketch map
e. Digitize the uncertain location points from the sketch map.

The uncertain location in this case, are the locations whose names are not know by the Namibian Mapping Agency. Attempt to get more information on these locations were made through the University of Namibia, History department, The Namibian Scientific Society (data source), Ministry of Lands and Resettlement and the Namibian Statistics Agency. Only map sketches were acquired to provide information on the locations of these place names. The objective is to handle the uncertain locations and be able to visualize them and to do that, the location points were digitized on a sketch map as shown in Figure 32: Digitising uncertain locations, as explained above. The map with black dotted symbols is the map sketch for an area in the Hereroland. The red dots are the uncertain locations digitized from the sketch map as explained above.

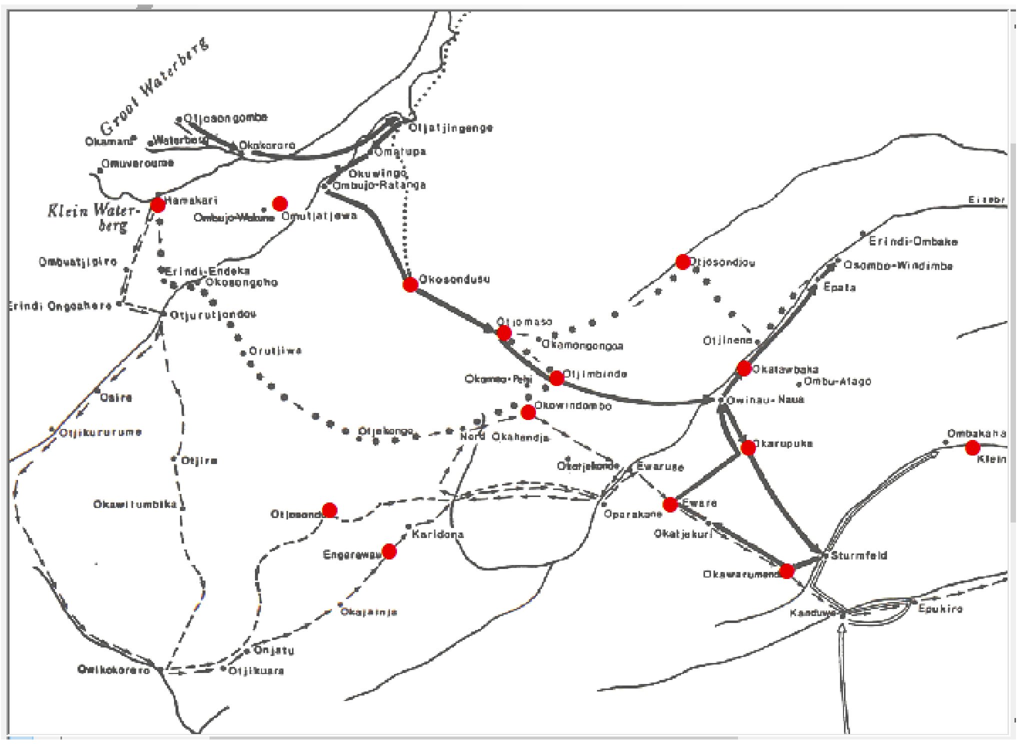


Figure 32: Digitising uncertain locations

5.2.1.2 Approximated location

It is very common in history to find location indicators that refers to the geographic location of an event. For instance, “They were lined up along the banks of Swakop River”. In a bigger picture, this is accurate in storytelling. One can create a mind image on the extent of the location and event happening. However, resolving such location description to exact geographic location is another challenge. At this point, there are no map sketches indicating the exact locations and therefore other measures need to be taken to define the location. Because locations are mapped as points in this research, the centre point of the river bank is assigned as the location depending on the direction of the next event in the chronological order as extracted.

5.2.1.3 Uncertain geographic positions.

Overall, location information in history do not give the exact geographic position of an event. For instance “The Hereros arrived in Omaruru”, this gives a location but not the geographic location. It could be that they were at the outskirts, in the centre or just close to Omaruru and the exact geographic position stays uncertain. For this reason this type of uncertainty is ignored in this research. The geographic position which is referred to by the location name is confined to centre point of the place. However, such uncertainties may not be acceptable for large scale and detailed mapping.

5.2.2 Temporal uncertainties

Temporal uncertainties comes from heterogeneous granularity that comes with the way the information is presented in the publications. Some events are associated to specific dates and other may be linked to ranges of dates. For example, *"Between July 13 and July 20, Chief Tjetjo met unit Volkmann in Karibib and violence erupted"*. There is an uncertainty on the exact date they met. The information extraction method used in our data, retrieves the data in a temporal order. This allows us to establish relations between individual troop's location visits events. The location visit of a troop for example with an unknown date, which falls between two know location visits dates is linked to the range determined by the two known location visit dates. The distance between the location of the unknown date and the subsequent known location visit date indicates the possibility of movement in days, which helps determine an approximate date. This is because there is no knowledge on the routes taken and mode of transportation.

Overall, modelling historical uncertainties is another research concept which entails indepth studying and implementation of methods to handle the inaccuracies. However, this is out of scope for the scope of this research.

CHAPTER 6: CONCLUSION AND RECOMMENDATIONS

6.1 CONCLUSIONS

Historical information provides us with the opportunity to study the past. With the growing developments in spatio-temporal information retrieval, exploring different data sources has become an easy task. The information extraction results demonstrate that we are no longer confined to traditional data retrieval methods in order to model the world around us.

Space-time analysis gives us an opportunity to answer spatio-temporal questions. However, the performance of a space-time analysis highly depends on the temporal and spatial granularity of the data presented. In this research, we used the space-time cube analysis method to identify the locations where the troops were concentrated, the time periods with more movements. We used spatio-temporal clustering to analyse and visualize the spatio-temporal relationship between the movements and battle events. This helps us understand why there were more movements at some locations compared to others. The battle events could have triggered some movements. With these, we were able to assess the ability of ArcGIS tools and functions in analysing and visualizing time in regards to discrete and poor temporal patterns. The results show that the analysis tools in ArcGIS performs well with multivariate, continuous temporal patterns and big amount of data. For example, the space-time cube tool requires a substantial amount of data to aggregate them into space-time cube bins. That is the reason why the individual trajectories could not be modelled using this tool. Another interesting tool is the grouping analysis functions that creates spatio-temporal clusters. However, its functionality requires more spatial and temporal occurrences in the data which was not the case with our data.

In regards to visualization, there are 2D and 3D visualization methods that facilitate the representation and display of temporal data. ArcGIS provides time animation functionalities to handle temporal overlaps when visualizing in 2D. The advantage of time animations is that they make it easier for viewers to detect changes in space and time. However, more functionalities to deal with moving object are still lacking.

6.2 RECOMMENDATIONS

The functionalities and capabilities of handling spatio-temporal data in GIS applications is not overlooked. We witness the growing developments and new ideas in combating issues regarding temporal information representation. New approaches and tools have been developed lately and specifically in ArcGIS application. However, more improvements are needed to see the realization of temporal GIS applications. To extend on to the evaluation section, future work need to be directed towards development of functionalities that exhaust the temporal aspect of the spatial information.

Imagine having built in functionalities that allows to query time in term of past, present and future state of an object. Questions such as “Where was an object two day before it reached a certain location” would be easier to answer.

In regards to analysis and representation of spatio-temporal data using space-time cube, having an option to query and manipulate the space-time cube contents would provide users with more information regarding the represented information in the cubes. Implementation of a space-time path functionality in ArcGIS that models trajectories in regardless of the amount of data would be very helpful. This would be useful in modelling historical trajectories that have dynamic temporal information.

It is difficult to represent moving objects in ArcGIS due to the lack of functionalities to estimate time and locations between known points. The display of moving paths as lines is unnatural and does not give a realistic representation. Therefore, having functionalities that shows moving progression would be useful in display moving objects.

Lastly, in regards to time animation and time slider functionalities, more improvements are needed to provide more time setting options. Such as allowing the display of objects at time of existence only, this can be very helpful in displaying important events. This is because time step concept displays objects that falls within the specified time step at once. Although this can be useful when visualizing multiple time layers and big amounts of temporal data, it does not work with dynamic discrete data in small amounts such as the historical battle events in this research.

REFERENCES

Bekele, M. K. (2014, March). *Spatial tracing of historic expedition: from hisroty to trajectory*. University of Twente.

Bley, H. (1971). *South-West Africa under German Rule 1894-1914*.

Bridgman, J. M. (1981). *The revolt of the Hereros*.

Dempsey, C. (2014, August 10). What is the Difference Between a Heat Map and a Hot Spot Map? Retrieved September 21, 2017, from https://www.gislounge.com/difference-heat-map-hot-spot-map/

Drechsler, H. (1966). *Let us die fignthing*.

Frihida, A., Zheni, D., Ghezala, H. B., & Claramunt, C. (2009a). Modeling Trajectories: A Spatio-Temporal Data Type Approach. In *2009 20th International Workshop on Database and Expert Systems Application* (pp. 447–451). https://doi.org/10.1109/DEXA.2009.70

Frihida, A., Zheni, D., Ghezala, H. B., & Claramunt, C. (2009b). Modeling Trajectories: A Spatio-Temporal Data Type Approach. In *2009 20th International Workshop on Database and Expert Systems Application* (pp. 447–451). https://doi.org/10.1109/DEXA.2009.70

Frihida, A., Zheni, D., Ghezala, H. B., & Claramunt, C. (2009c). Modeling Trajectories: A Spatio-Temporal Data Type Approach. In *2009 20th International Workshop on Database and Expert Systems Application* (pp. 447–451). https://doi.org/10.1109/DEXA.2009.70

Gregory, I. (2002). A Place in History - Section 1.3. Retrieved September 26, 2017, from http://hds.essex.ac.uk/g2gp/gis/sect13.asp

Historical method. (2017, August 21). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Historical_method&oldid=796597222

Jacquez, G. M. (2008). Spatial cluster analysis. *The Handbook of Geographic Information Science*, *395*, 416.

Janković, S., Mladenović, S., Uzelac, A., Zdravković, S., & Aćimović, S. (2017). Development of Traffic Geo-Application Based on Big Data Processing (pp. 458–463). Singidunum University. https://doi.org/10.15308/Sinteza-2017-458-463

jpeacock. (2016, July 14). A Brief History of Namibia: Deutsch Südwestafrika. Retrieved October 5, 2017, from https://peacockofthekalahari.wordpress.com/2016/07/14/a-brief-history-of-namibia-deutsch-sudwestafrika/

Katz, P., & Schill, A. (2013). To learn or to rule: two approaches for extracting geographical information from unstructured text. *Data Mining and Analytics 2013 (AusDM'13)*, 117.

Klaus, D. (2000). 38-1884. Retrieved October 5, 2017, from http://www.klausdierks.com/Chronology/38.htm

Kordjamshidi, P., Frasconi, P., Otterlo, M. V., Moens, M.-F., & Raedt, L. D. (2011). Relational Learning for Spatial Relation Extraction from Natural Language. In *Inductive Logic Programming* (pp. 204–220). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31951-8_20

Kossatz, M. (2016a). Visualization an editing the history of countries in time and space with Histoglobe.

Kossatz, M. (2016b, June 10). *Visualizing and editing the history of countries in time and spce with HistoGlobe* (Master). Bauhaus-Universitet Weimar, Weimar.

Kraak. (2003). Cartographic renaissance: 21st International Cartographic Conference, 10-16 August 2003, ICC, Durban. Document Transformation Technologies.

Kwan:, M.-P. (2007). Mei-Po Kwan: Color ../TGIS/Figures of Kwan's TRC paper. Retrieved September 30, 2017, from http://meipokwan.org/TGIS.htm

Li, X. (2005). *New methods of visualization of multivariate spatio-temporal data: PCP time-cube and multivariate time-cube*. University of Twente. Retrieved from http://www.itc.nl/library/papers_2005/msc/gfm/xia.pdf

Nadi, S., & Delavar, M. R. (2003). Spatio-Temporal Modeling of Dynamic Phenomena in GIS. In *ScanGIS* (pp. 215–225). Retrieved from http://ai2-s2-pdfs.s3.amazonaws.com/f365/32c6bda910059eac9a14ca4fc1130386de06.pdf

Namibia 1on1. (2013). Herero Uprising Namibia. Retrieved October 5, 2017, from http://www.namibia-1on1.com/herero-uprising.html

Owens, J. B. J. (2007). What Historians Want from GIS. Retrieved October 7, 2017, from http://www.esri.com/news/arcnews/summer07articles/what-historians-want.html

Pelekis, N., Theodoulidis, B., Kopanakis, I., & Theodoridis, Y. (2004). Literature review of spatio-temporal database models. *The Knowledge Engineering Review*, *19*(3), 235–274.

PG-Trajectory.pdf. (n.d.). Retrieved from http://pg-trajectory.dmlab.cs.gsu.edu/documents/PG-Trajectory.pdf

Rattenbury, T., Good, N., & Naaman, M. (2007). Towards automatic extraction of event and place semantics from flickr tags. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 103–110). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1277762

Shaw, S.-L. (2006). Towards a GIS-based Analytical Time-geographic Framework for Physical and Virtual Activities. Retrieved September 30, 2017, from http://web.utk.edu/~sshaw/NSF-Project-Website/default.htm

Shaw, S.-L., Yu, H., & S. Bombom, L. (2008). A Space-Time GIS Approach to Exploring Large Individual-based Spatiotemporal Datasets. *T. GIS*, *12*, 425–441. https://doi.org/10.1111/j.1467-9671.2008.01114.x

Strötgen, J., Gertz, M., & Popov, P. (2010). Extraction and exploration of spatio-temporal information in documents (p. 1). ACM Press. https://doi.org/10.1145/1722080.1722101

Thakker, D., Osman, T., & Lakin, P. (2009). Gate jape grammar tutorial. *Nottingham Trent University, UK, Phil Lakin, UK, Version, 1*. Retrieved from https://gate.ac.uk/sale/thakker-jape-tutorial/GATE%20JAPE%20manual.pdf

Wang, W. (2014). *Automated spatiotemporal and semantic information extraction for hazards*. The University of Iowa. Retrieved from http://search.proquest.com/openview/07e9131edc0a76b2899465a4b3daf918/1?pq-origsite=gscholar&cbl=18750&diss=y

Wang, X., Zhou, X., & Lu, S. (2000). Spatiotemporal data modelling and management: a survey. In *Technology of Object-Oriented Languages and Systems, 2000. TOOLS-Asia 2000. Proceedings. 36th International Conference on* (pp. 202–211). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/885919/

Wikipedia. (2017, September 20). Time. In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Time&oldid=801572100

APPENDIX

Appendix A: Temporal transducers

```
Phase:LocationFinder
Input:Location Lookup SpaceToken Token
Options: control = appelt

//Macro
Macro: NE
({Lookup.majorType=="direction"})

Macro:KM
(	({SpaceToken})?
	({Token.kind==number,Token.length=="1"}|
	{Token.kind==number,Token.length=="2"}|
	{Token.kind==number,Token.length=="3"}|
	{Token.kind==number,Token.length=="4"}
	)(
	({SpaceToken})?
	{Token.kind==word,Token.category==NN,Token.orth==lowercase,Token.string=="
km"}|{Token.string=="miles"}
	)

)

/////Between Windhoek and Karibib
Rule:Location1
priority: 50
(
	({SpaceToken})?
	({Token.string == "between"}|{Token.string == "Between"})
	({SpaceToken})?
	({Location})
	({SpaceToken})
	({Token.string == "and"})
	({SpaceToken})
	({Location})

):Location1
-->
	:Location1.SpatialRelation = {rule= "Location1"}

////////North of Windhoek
Rule: Location2
Priority: 50
(
	({SpaceToken})
	({Lookup.majorType=="direction"})
```

```
        ({SpaceToken})
        ({Token.string=="of"})
        ({SpaceToken})
        ({Location})
):Location2
-->
        :Location2.SpatialRelation = {rule = "Location2"}

/////Near Windhoek
Rule:Location3
Priority: 40
(
        ({Token.string=="near"})
        ({SpaceToken})?
        ({Location})
):Location3
-->
        :Location3.SpatialRelation = {rule= "Location3"}

/////Area of Okahao, Oshakati and windhoek
Rule:Location4
Priority: 40
(
        ({Token.string== "area"})
        ({SpaceToken})?
        ({Token.string== "of"})
        ({SpaceToken})?
        ({Location})
        ({Token.string== ","})
        ({SpaceToken})?
        ({Location})
        ({SpaceToken})?
        ({Token.string== "and"})
        ({SpaceToken})?
        ({Location})
)
:Location4
-->
        :Location4.SpatialRelation = {rule = "Location4"}

//////Area of Oshakati and Ongwediva
Rule:Location5
Priority: 40
(
        ({Token.string== "area"})
        ({SpaceToken})?
        ({Token.string== "of"})
        ({SpaceToken})?
        ({Location})
        ({SpaceToken})?
```

```
    ({Token.string== "and"})
    ({SpaceToken})?
    ({Location})
)
:Location5
-->
    :Location5.SpatialRelation = {rule = "Location5"}

//////20km north west of Omangeti
Rule:Location6
Priority: 50
(
    (KM)
    ({SpaceToken})
    ({Lookup.majorType=="direction"})
    ({SpaceToken})
    ({Lookup.majorType=="direction"})
    ({SpaceToken})
    ({Token.string=="of"})
    ({SpaceToken})
    ({Location})

):Location6
-->
    :Location6.SpatialRelation = {rule = "Location6"}

//////20km north of Omangeti
Rule:Location7
Priority: 50
(
    (KM)
    ({SpaceToken})?
    ({Lookup.majorType=="direction"})
    ({SpaceToken})?
    ({Token.string=="of"})
    ({SpaceToken})?
    ({Location})

):Location7
-->
    :Location7.SpatialRelation = {rule = "Location7"}

//////areas north-east of Omangeti
Rule:Location8
Priority: 40
(
    ({Token.string=="areas"})
    ({SpaceToken})?
    ({Lookup.minorType=="direction"})
    ({SpaceToken})?
```

```
	({Token.string=="of"})
	({SpaceToken})?
	({Location})

):Location8
-->
	:Location8.SpatialRelation = {rule = "Location8"}


///south west of Okahandja
Rule:Location14
Priority:50
(
	({Lookup.majorType=="direction"})
	({SpaceToken})
	({Lookup.majorType=="direction"})
	({SpaceToken})
	({Token.string=="of"})
	({SpaceToken})
	({Location})

):Location14
-->
	:Location14.SpatialRelation = {rule ="Location14"}


////////North of the Windhoek
Rule: Location9
Priority: 50
(
	({SpaceToken})
	({Lookup.majorType=="direction"})
	({SpaceToken})
	({Token.string=="of"})?
	({SpaceToken})?
	({Token.string=="the"})?
	({SpaceToken})?
	({Location})
):Location9
-->
	:Location9.SpatialRelation = {rule = "Location9"}

//////from Otjimanangombe via Epata, Otjosondu and Osondema to Otjituuo - NOT WORKIN
Rule:Location10
Priority: 40
(
	({Token.string=="via"})
	({SpaceToken})?
	({Location})
```

```
):Location10
-->
	:Location10.SpatialRelation = {rule = "Location10"}

/////between the little and great Waterberg
Rule:Location11
Priority:40
(
	({Token.string == "between"})
	({SpaceToken})?
	({Token.string == "the"})
	({SpaceToken})?
	({Token.string == "Little"})
	({SpaceToken})?
	({Token.string == "and"})
	({SpaceToken})?
	({Token.string == "Great"}|{Token.string == "Mount"})
	({SpaceToken})?
	({Location})
):Location11
-->
	:Location11.SpatialRelation = {rule = "Location11"}

Rule: Location12
Priority: 50
(
	({Token.string=="to"})
	({SpaceToken})
	({Token.string=="the"})
	({SpaceToken})
	({Lookup.majorType=="direction"})
	({SpaceToken})
	({Token.string=="of"})
	({SpaceToken})
	({Location})
):Location12
-->
	:Location12.SpatialRelation = {rule = "Location12"}

/////4km downstream of Gross Barmen
Rule:Location13
Priority:50
(
	(KM)
	({SpaceToken})
	({Token.string=="downstream"})
	({SpaceToken})
	({Token.string=="of"})
	({SpaceToken})
	({Location})
```

```
):Location13
-->
	:Location13.SpatialRelation = {rule= "Location13"}


//// about 25 miles.
Rule:Location14
(
	({Token.string=="about"})
	({SpaceToken})
	({Token.kind==number,Token.length=="2"}|{Token.kind==number,Token.length=="
3"})
	({SpaceToken})
	({Token.string=="miles"})
):Location14
-->
	:Location14.SpatialRelation = {rule ="Location14"}
```