UNIVERSITY OF THE WITWATERSRAND

SCHOOL OF GEOGRAPHY, ARCHAEOLOGY AND ENVIRONMENTAL SCIENCES

GIS & REMOTE SENSING

RESEARCH DISSERTATION GEOG7000

Student Name	Bellina Chokoe	Student No.	1727895
Signature	Degree	MSc GIS & RS	
Email	1727895@students.wits.ac.za	Contact	0658074236
Research Title	Assessing the Suitability of the South African Rock Art Digital Archive for the Purpose of geocoding Rock Art Sites		
Supervisor	Samy Katumba		

Bellina Machuene Chokoe

A Master's research project submitted to the Faculty of Science, University of the Witwatersrand, in partial fulfilment of the requirements for the degree of Master of Science in GIS and Remote Sensing. Johannesburg, 2021.

Declaration

I, Bellina Machuene Chokoe, declare that this Research Report is my own, unaided work. It is being submitted for the Degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

21 April 2021

Abstract

The South African Rock Art Digital Archive (SARADA) is a repository of memory and a testament to over 30.000 years of human activities including interactions between people and the environment across the African continent. This archive is of great value to the African history and cultural depictions. However, incomplete information related to suitable geographic coordinates present a challenge for locating and accessing a considerable number of rock art sites, in this case considering the 780 rock art sites without geographic coordinates. This study assesses information records of rock art sites contained in SARADA to extract suitable textual information that can be used for the purpose of geocoding rock art sites. Geocoding methods are useful in transforming textual information into geographic coordinate locations. In order to determine whether the extracted textual information from SARADA was suitable for the purpose geocoding, attribute accuracy and completeness were assessed. Three standard geocoding methods were explored: Geocoding using addresses, geocoding using postal codes and geocoding using administrative boundaries. The textual information about rock art sites extracted from SARADA only allowed the geocoding of rock art sites to be performed at the postal code level (i.e. zip codes as geographic identifiers). The geocoding match rate is 62% which geocoded at a road level.

Acknowledgement

I would like to appreciate my supervisor Samy Katumba for taking on my project in such a short notice and guiding me until submission. I would like to thank Dr Stefania Merlo for being very strict and challenging me in this research, I have learned very much from her determination in the Geographical Information Systems field. Most importantly I would like to thank my family for their emotional support throughout my journey, it would not be possible without each and every one of them. To my big sisters Rebecca Buiswa Chokoe and Kwena Lamola, you have believed in me from the beginning. To my friend Elvis Malope, your intelligence and guidance is the reason I have made it this far. Lastly, I would like to wholeheartedly thank myself for not giving up when it seemed impossible.

List of Abbreviations

GIS: Geographical Information Systems

SARADA: South African Rock Art Archive

RARI: Rock Art Research Institute

STATS_SA: Statistics South Africa

ESRI: Environmental Systems Research Institute

SAPO: South African Postal Office

Table of Contents

List of Figures

Table of Figures

1.Introduction

1.1.Background

Rock paintings and engravings are Africa’s oldest continuously practiced art form. The form, function, and spatial distribution of rock art can reveal a great deal of information about the actions of the people, their environment, and the system in which they are deeply embedded (Chippindale and Nash, 2004). A rock art constitutes a repository of memory and a testament to over 30.000 years of human activity including interactions between people and the environment (Haubt, 2015).

A digital or analogous (i.e. on a drawing or paper map) rock art record can be used in a number of ways by different audiences. Campbell (2013) states that custodians require a precise, detailed and measurable record of the rock art, its condition, and its surroundings in order to monitor change and make informed decisions on management and conservation. According to Hall et al. (2016) researchers want a wide range of information on the content and context of the rock art. Interested members of the general public want a clear and accurate record of where the rock art is and what it looks like (Merryman, 1989). A wider audience, including learners and the public, want a visually exciting, accessible and engaging record that stimulates their imagination and learning experience (Hall et al., 2016).

Accessing the rock art information in the form of spatial data (e.g. geographic coordinate locations) provides opportunities for mapping and visualising the interaction between people and their geographical and cultural surroundings across space and time.

The South African Rock Art Digital Archive (SARADA) is one of the largest rock art digital archive in the world with collections from more than 30 institutions and individuals, giving access to more than 270,000 digital images of rock art. SARADA is available on a digital platform created by the Rock Art Research Institute (RARI). RARI holds a significant collection of historical documents, photographs, redrawings and slides in addition to its large working collection of slides, tracings and redrawings available on the South African Rock Art Digital Archive.

However, the current data structure of SARADA is stored in a non-spatial repository and does not support spatial searches or queries. Furthermore, data accessibility is limited to selected metadata and non-geo-referenced imagery of the rock art information.

Geocoding is a method that will be used in this research to convert existing suitable textual information of rock art sites into geographic coordinates of non-georeferenced rock art sites. The term geocoding refers to the process of coding the earth with geographic reference information that can be reconciled for computer mapping uses (Harries,1999). Geocoding involves transforming indirect spatial information into direct locations which can be placed on a map, thus allowing for spatial analysis and cross-referencing (Sanchez and Bertolotto, 2017). According to Kellison (2012) geocoding provides one approach to spatialising street addresses, parcels, or administrative zones.

1.2 Problem statement

Spatial information contained in SARADA is limited and not structured within a GIS-based system. Such information includes descriptions of the landscape where the rock art is found, the demographics and the human history as well as the positioning of the rock art tradition in time. However, not all rock art sites have geographic coordinate locations. This makes it a challenge for SARADA's researchers and potential users to perform spatial queries or spatial proximity analyses using the digital rock art archive as it is in its current format. This ultimately hinders efforts towards a thorough understanding of the richness and depth of information that one can extract from such a historic archive as alluded in the background section (section 1.1) of this research report. Geocoding is an inevitable step in any spatially based study with considerable bodies of data. This makes it a critical process in various contexts such as historical geocoding. Many geocoding services have been developed to fulfil this need, originating from private initiatives (Google Geocoding API,Mapzen (mapzen.com)), public agencies or from the open-source community (OpenStreetMap Nominatim (nominatim.openstreetmap.org), Gisgraphy (gisgraphy.com)). (Healy and Gilliland, 2012, Ng et al., 1993, Zandbergen, 2008) have provided enormous contribution on assessing geocoding tools using historical data. Searching SARADA for motifs or pictographs by location selection results in unrequested results. Furthermore, determining the proximity of ancient species in a desired area results in null responses as the spatial attribute is not encompassed in the database for queries.

1.3 Aim

The aim of this research project is to assess the suitability of the textual information extracted from SARADA for the purpose of geocoding rock art sites.

1.4 Motivation

Rock art sites are amongst the most vulnerable heritage sites due to the impact of climate change, expansion of urban areas and vandalism (UNESCO World Heritage Centre, 2010). It is becoming important to implement conservation and preservation methods as more and more rock art sites are being lost due to natural and cultural threats. Besides serving as a digital repository, extracting spatial data could assist in improving an intelligent retrieval of rock art information through spatial queries and proximity analyses. For example, rock art sites could be queried in relation to other physical and natural features that are in the vicinity of such sites. This will improve the understanding of rock art by researchers and the general public. For this to happen, all rock sites recorded in SARADA will need to be georeferenced. At present, besides having geographic coordinate locations of some rock art sites, the spatial information contained in the SARADA is in a textual form (e.g. site name, place name etc.). Therefore, the main

1.5 Research Objectives:

The objectives of the project are as follows:

1. Identify and assess suitable textual information that can be used to extract spatial

information for the purpose of geocoding rock art sites.

2. Design a geocoding routine and perform geocoding and accuracy assessment based on the extracted spatial information obtained in the first objective.

2. Literature review

Over the years, the changing availability of geographic data has forced the concept of geocoding to remain flexible and adaptive in terms of its requirements and capabilities (Vieira et al, 2010). The increasing availability, accuracy, and reliability of digital geographic reference datasets has meant that the geocoding process has continually evolved to keep pace with the underlying datasets that facilitate its use (Vieira et al, 2010). More modern attempts at geocoding have tackled the problems of assigning valid geographic codes to far more types of locational descriptions such as street intersections (Levine and Kim 1998), enumeration districts (census delineations) (Sheehan et al. 2000), postal codes (zip codes) (Gatrell 1989, Collins et al. 1998, Sheehan et al. 2000, Krieger et al. 2002b, Hurley et al. 2003), named geographic features (Davis et al. 2003, United Nations Economic Commission 2005), and even freeform textual descriptions of locations (Wieczorek et al. 2004, Hutchinson and Veenendall, 2005).

2.1 Geocoding archival of non-address data

The geocoding of cultural content is becoming a very promising technique which is introducing new scenarios of exploring and improving heritage sites across the globe (Vodeb and Zakrajsek, 2013). Guidelines and recommendations were established by Kollen et al. (2013) to provide an access to geographic (spatial) data and tools as well as how to catalogue non-address data. Several applications have been established and are used in the cultural tourism, teaching, research domains and e-infrastructure using GIS (Zakrajsek and Vodeb 2013). Kollen et al. (2013) investigated techniques of storing spatial data and made recommendations on the technology and software needs for transforming historical data to digital content. The findings provided information on the current geospatial data catalogue practices and approaches by academic libraries and the conclusion was that users can conduct a wide range of research as well as planning, analysis, and decision making.

Nevertheless, there are still quite some challenges in terms of geocoding archival of non-address data. For example, seldom studies have used archival classification standards in order to define and catalogue the underlying geographical historical data and information (Titus, 2016). Both Sanchez and Bertolotto (2017) and Cura et al., (2018) mention that historical data is filled with uncertainties (pertaining to temporal, textual and positional accuracy, as well as to the reliability of historical sources) which can either be ignored or entirely be resolved.

Geocoding tools that work efficiently for current addresses such as Bing Maps, ESRI address locator and USC geocoding platform were assessed by (Swift et al., 2008). However, these do not tackle temporal information, and usually follow a very strict hierarchy (country, city, street, house number, etc.) which is very challenging, if not impossible, to use with historical data (e.g. rock art).

2.2 Geocoding Methods

Geocoding involves the process of assigning geographic coordinates (latitude and longitude) to a location (or a place) by simply comparing the descriptive location-specific elements against a reference geospatial dataset (Zandberg, 2008). ESRI (2014) outlines the geocoding process steps involved in translating an address entry, searching for the address in the reference data, and output the best match candidate or candidates as a point feature on the map. Dao (2015) and Kellison (2012) outline geocoding steps which include the following:

- Parsing the input address into address components (such as street name, street type, etc.) as a first category,
- Standardizing abbreviated values,
- Assigning each address element to a category known as a match key, indexing the needed categories, searching the reference data,
- Assigning a score to each potential candidate, filtering the list of candidates based on the minimum match score and delivering the best match.

Research on geocoding methods by Kellison (2012), Titus (2016) and Zandbergen (2008) demonstrates that even though geocoding applications vary and span numerous types of applications, there are several common problems that were identified. For example, poor matching rates, requiring excessive manual mapping by the user, inaccuracies and/or incompleteness of the resulting spatial datasets. It should be noted that spatial datasets have become an essential element in geographic information system (Goodchild, 2003). In the context of the study presented in this report, the process of geocoding results in the creating of spatial datasets of rock art sites.

2.2.1 Geocoding by Addresses

In 2004, Standards South Africa, a division of the South African Bureau of Standards (SABS), began a project of developing a South African National Standard (SANS) for 'a standard framework for South African addresses', subsequently given the designation SANS1883 (Coetzee and Cooper, 2007). The aim of the standard was not to devise a new system of addressing or to build a national address database, but rather according to Van Rensburg (2015), this was to enable interoperability in address data sets and geographical information systems (GISs), which will facilitate developing a national address database.

The information captured in SARADA does not follow the standard framework for south African addresses. The web-database does not provide site location data other than the geographic coordinate locations (latitude and longitude) of some rock art sites. However, SARADA contains imageries organized by institution or contributor. This is one of the challenges for geocoding rock art data. According to Coetzee and Cooper (2007) formal addresses have never been assigned in vast areas of South Africa. These areas include farms, rural villages and former black townships. In the SARADA database, directions or addresses are given in textual forms such as "How to find site" whereby it highlights the direction or paths to take in order to get to a specific rock art site.

2.2.2 Geocoding by Postal codes

The postal code system of South Africa was launched on the 8th of October, 1973 (Rossow, 2008). A postal code is a 4-digit code that relates to the region, specific to post office distribution, postal sorting centre (HUB), and to some extent, province (Lombaard, 2010). It is ordered numerically according to a distribution area and as such, the sorting ranges do not directly relate to province. Postal codes are often the only geographic identifier available for assigning contextual or environmental information to a study population (Charif et al., 2010).

The accuracy of geocoding from postal codes can vary and geocoding imprecision may result in misclassification, depending on the spatial resolution of the environmental or contextual measures. According to Charif et al (2010) and Tian et al (2016) geocoding based on a full street address is a highly accurate way of assigning geographic coordinates to an individual's residential location than using postal codes. However, the SARADA address information is not standardized and fully available in this format.

2.2.3 Geocoding by boundaries

Geocoding by boundary is the least accurate of the three methods of mapping point locations (Dramowicz, 2004). Any boundary file, such as provinces or districts, can be used, as long as the boundary name in the table to be geocoded can be matched with the boundary name in the reference table. The larger the boundary, the less accurate the results (Charif et al., 2010, Dao, 2015, Dramowicz, 2004). The three standard geocoding methods are explored in order to transform non-spatial data to spatial data.

2.3 Challenges on geocoding historical data

Sanchez and Bertolotto (2017) developed a historical GIS database for the City of Dublin (Ireland) and in developing a system for storing and manipulating historical data related to the visitors of a library in the city of Dublin in the time period 1826-1926, they faced several challenges linked to the uncertainty associated with such data. According to Sanchez and Bertolotto(2017) these were due to several different aspects including lack of consistency in the recording of visits to the library over time, interpretation of different handwriting styles, varying levels of granularity in both the spatial (address of visitors) and temporal (date of visit) dimension of the records.

The approach they took was inspired by the work of (Malizia, 2013) who defines three types of inaccuracies as locational, incompleteness and temporal which are the biggest challenging factors in assessing historical GIS data.

A geocoding process of a set of addresses will be more successful if the percentage of addresses that can be associated with geographic coordinates is high. To achieve this high percentage, a loosening of conditions may be required to accommodate some trade-offs between the rate of matches and locational accuracy. Therefore, the positional accuracy of those coordinates will likely decrease (Sanchez and Bertolotto, 2017). Malizia (2013) mentions three reasons which lead to poor geocoding outcomes: misspelled or abbreviated addresses; records which include alternative identifiers to address (e.g. PO box numbers); or missing attribute information in the reference file.

Van Rensburg (2015) conducted a study in South Africa to determine a geocoding workflow for South African Data and concluded that South African addresses are not standardised. This was also previously mentioned by Coetzee and Cooper (2007).

2.4 Geocoding match rates

The simplest measure of geocoding quality is the match rate, or the percentage of records that produce a reliable match (Tian et al., 2016, Zandbergen, 2008). An obvious question that emerged in most geocoding literature is: What is an acceptable match rate? But this question has been given limited attention in the literature. In general, match rates reported in studies that have used geocoding vary greatly since they depend on many factors (Dao, 2015, Kellison, 2012, Tian et al., 2016, Zandbergen, 2008). There is no agreement on a common standard for an acceptable geocoding match rate.

According to (Cura et al., 2018) and (Vodeb and Zakrajsek, 2013) the match rates increases if efforts are made to increase the quality of the address file and the geographic reference file. Interpreting match rates, however, is very subjective since much rely on the criteria used to characterize a match. The match rate is based on how well the locations found in the reference data match with the tested data.

The match rates are based on a weighted numbering system; based on the number of matching characters in each of the prioritized/configured address element areas as well as zip codes. So, the more characters that can match the better the likelihood of a high score. Every individual character contributes a percentage to the total score. This is based on the algorithm of address locators in ESRI ArcGIS. When using postal/zip code boundaries, a match rate goes to the correct 4-digit code. Scores are weight summed, with percentage normalization. Missing elements do not penalize a score, they simply do not contribute (Zhan et al., 2006) and (Zandbergen, 2008).

2.5 Geographical Data Quality Assessment

Given that the process of geocoding rock art sites recorded in SARADA would result in the creation of geospatial data (i.e. geographic coordinate locations of rock art sites), it is relevant to review how the quality of geospatial datasets are assessed. Prescott (1996) comprehensively outlines quality as an essential or distinguishing characteristic necessary for the spatial or geographic data to be fit-for-use; meeting an expectation; degree of excellence and conformance to a standard. Furthermore, quality is about meeting satisfaction with a data product. A satisfactory level is commonly met by “conformance to a standard” or a specification, i.e. by a specified set of regulations for a data producer to follow in order to meet an agreeable quality level (Prescott, 1996).

2.5.1 Components of data quality assessment

Accuracy is defined as the degree or closeness to which the information on a map matches the values in the real world (Veregin, 1999). However, spatial datasets are subject to a variety of quality issues, due to factors related to data acquisition and calibration (Levesque et al., 2007). For example, assessing the quality of the data in terms of its attribute accuracy, can help to assure its suitability within a specific application area (Fischer and Getis, 2009).

2.5.1.1 Positional Accuracy

Positional accuracy is basically related with ensuring that lines and polygons have the correct shape and are correctly located (Docan, 2013), can also be defined by the accuracy of positions of spatial features of a data set in correspondence to the geographical space. Positional accuracy is the accuracy of positions of spatial features depicted within a spatial data set in relation to a prevailing projected geodetic reference system (Zandbergen, 2008). Furthermore, Zandbergen (2008) explains that “positional accuracy is basically related with ensuring that lines and polygons have the correct shape and are correctly located”. In this study, the positional accuracy was not assessed because there is no ground truth data from a reference file of rock art sites and the geocoding process is to determine positional points of the sites.

2.5.1.2 Attribute accuracy

Another relevant element of data quality is the attribute accuracy. The term attribute is referred to a fact or a specific property about some geographic feature or location (e.g. elevation, land cover class, etc.), stored as spatial information in a GIS database (Van Oort, 2006). According to Haklay (2010) Attribute accuracy is the closeness of attribute values to their true value. The attribute accuracy was assessed in this study to determine the producer and user accuracy that will further assist in determining if the geocoding match rate is affected by the data quality.

2.5.1.3 Completeness

Completeness in the simplest term shows if each entity occurrence is present and whether all its attributes are present. Entity occurrence can be defined as the real-world phenomenon of a given type. Completeness also refers to the presence and absence of spatial features, associated attributes, and relationships between features within a dataset (Siebritz, 2014). Completeness focuses on information of selection criteria, definitions and mapping rules (Docan, 2013). It is important in this study to understand the omission and commission error of the datasets, therefore the completeness of the dataset in this research was assessed.

2.5.1.4 Lineage

The lineage contains descriptions of data sources, methods which are used for databases designs, including all data transformations and transactions used in the development of data (Joksić and Bajat, 2004). This component must include all data significant for both, data sources and for the data upgrading process. Joksić and Bajat (2004) mention that "lineage is usually the initial component given in data quality reports, because all other spatial data quality components are subordinate on data lineage". The final purpose of lineage report is to keep valuable information of data history for future users and this can be seen as metadata. This assessment was compiled from the Rock Art Research Institute history and working along with the director of ringing rocks digitisation laboratory, the metadata of the database was compiled.

2.6 Geocoding tools assessment

There are numerous software tools being used to geocode data. The most commonly used tools are Google Maps, Google Earth Pro, ESRI ArcGIS and Bing Maps, these tools are compared and selected for evaluation as they are most commonly used geocoding applications. “Google, which is the most preferred on a global scale, has limitations on how the data can be displayed”. “According to the Maps API Terms of Service Licence restrictions”, the “Geocoding API may only be used in conjunction with a Google map” (Google, 2013).

Table 1: Assessed geocoding tools (Van Rensburg, 2015)

Geocoding Tool	Address Accuracy (South African)	Address Accuracy (South African)	Geocode Output	Usability	Cost	Additional Info	Used in Study
Bing Maps API	Street Centroid	Upload a CSV file to a browser based Geocode Dataflow API	Geographic Coordinate	Simple, but the connection often times out	Only free up to 10,000 Geocodes per 24 hour period	Requires a fee to register API key	NO
Centrus	Not for South Africa	N/A	Geographic Coordinate	N/A	N/A	NO	NO
Esri (World Geocode Service)	Point or Street Segment	DB connection or Excel/CSV upload to application with connection to central geocoding server	Geographic Coordinate	Simple. Both raw and mapped data available	Free: 2500 Geocodes; Entry level cost: R29,000 ~R1 a Geocode	Desktop Software requires licence /geocoding Subscription	**YES**
Gazetteer	Postal Centroid	Import flat files to a database; Match suburb and town	Geographic Coordinate	Time intensive, often produces duplicates	Free	Not always up to date	NO
Geocoder. us	Not for South Africa	N/A	Geographic Coordinate	N/A	N/A	NO	NO
Quantum GIS geocoding web service	Map markers/p	Excel/CSV upload to application with	Geographic Coordinate	N/A	N/A	Yes	NO

	oint markers	connection to central geocoding server					
Google Earth Pro	Building or Landmark	N/A	Geographic Coordinate	N/A	N/A	N/A	Only secondary
Google Maps API	Building or Landmark	JavaScript API	Geographic Coordinate on Google Maps only	Raw data is not available in bulk	Free: 100,000 requests per day	Requires free registered API key	Only secondary
MapQuest	Not for South Africa	N/A	Geographic Coordinate	N/A	N/A		NO
Yahoo Maps API	Service no longer operating	N/A	N/A	N/A	N/A		NO
USC Geocoding Platform	Not for South Africa	N/A	Geographic Coordinate	N/A	N/A		NO

The table above shows a distinctive evaluation of the tools to use in this research. Google, which has the most accurate results worldwide, has limitations on how the data can be displayed. As such, Google can be used only as a reference. Among the evaluated geocoders, ESRI geocoding service was used.

2.7 Geocoding Matching rate

The geocoding matching rate used in this study is adopted from (Tian et al. 2016) and (Zandbergen, 2008). The accuracy of geocoding is one of the biggest concerns highlighted from both studies (Tian et al. 2016; Zandbergen, 2008). The following criteria are used to determine the matching degree of geocoding results.

Table 2: Geocoding match rates

"Matching degree"	"Matching Level Information"
100%	"Matched with accuracy"
90-100%	"Matched at the building level (building numbers); building numbers interpolation is possible"
80-90%	"Matched at the block level"
70-80%	"Matched at the community level"
60-70%	"Matched at the road level"
50%-60%	"Matched at the sub-district level"
50%-0%	"Matched at the district level"
Unmatched	"Unmatched at any level"

According to (Tian et al., 2016, Zandbergen, 2008), a matching degree shows the quality of a geocoded record. A record is accurately matched, inaccurately matched or unmatched.

3. Research Methodology

3.1 Study Area

The South African Rock Art Digital Archive, which is now known as The African Digital Archive, includes information and depictions of rock art sites across the African continent. Majority of the information in the archive is geographically located in South Africa. The Limpopo province of South Africa has been identified as the study area of this research. The reason being, this province presents rock art diversity with a total number of 1180 of rock art sites recorded, some with and others without precise geographic coordinates. As such, this study area is ideal for assessing the suitability of the South African Rock Art Digital Archive for the purpose of geocoding Rock Art Sites. Limpopo is located in the northern part of South Africa and shares border with three countries: Botswana, Mozambique and Zimbabwe. This province has a great diversity of cultural heritage and is home to the World Heritage Site of Mapungubwe and the Stone Age and Iron Age relics of the Makapansgat Valley.

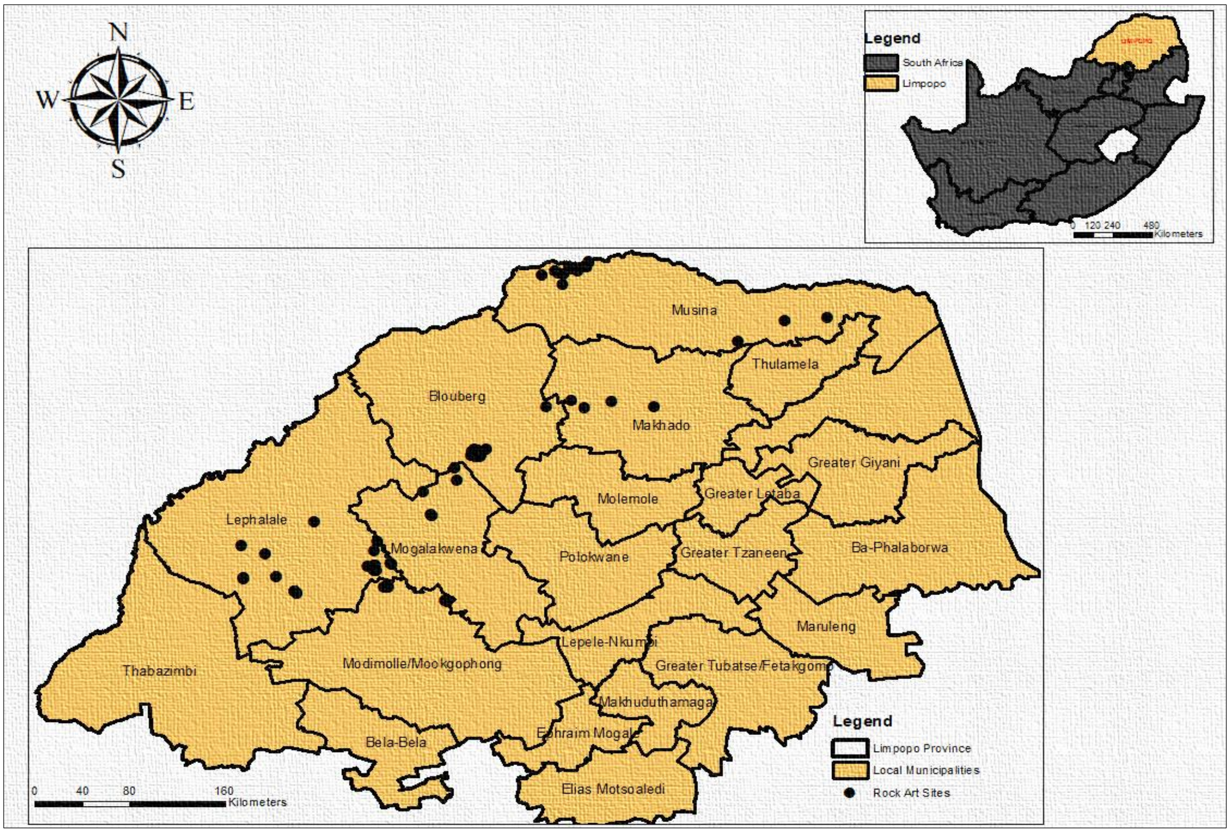


Figure 1 : Map of the Limpopo Province as the Study Area

3.2 Methods

This section presents the details of the procedures and tools that are employed to accomplish the objectives of this research. A discussion of the data requirements and geocoding techniques implemented is also provided.

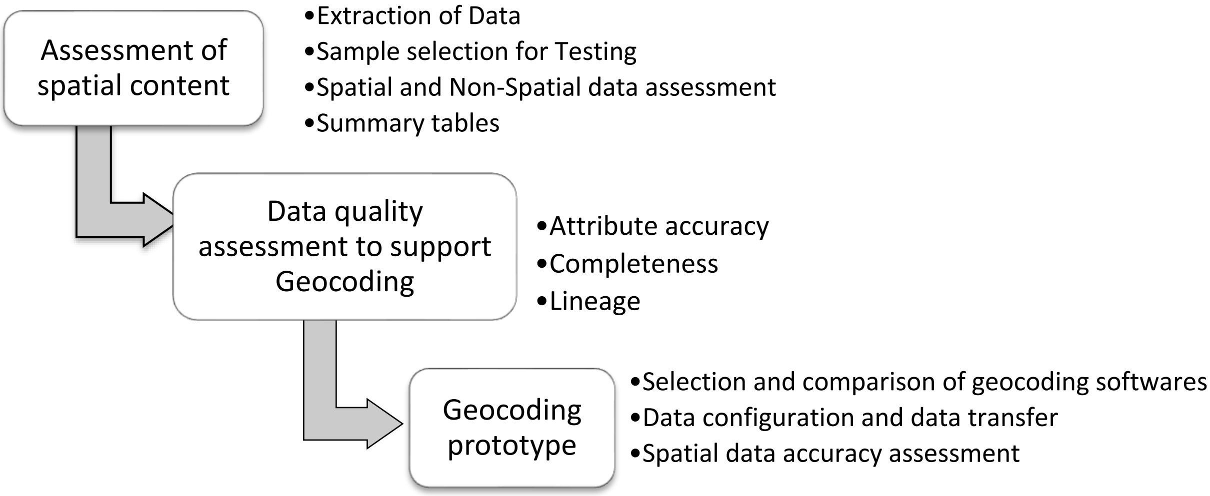


Figure 2 : Workflow showing geocoding methodology

The first process was to assess the spatial content of the database, both spatial and non-spatial attributes were extracted to a common file format and categorized into summary tables as seen from table 3 to table 5. ESRI ArcGIS is selected as a compatible geocoding software. The final step was assessing the spatial data accuracy following the data quality component as described in the literature review.

3.2.1 Assessing the spatial capabilities of SARADA

SARADA contains both spatial and non-spatial information although set up in an environment that is not capable of performing spatial queries. The information captured within this database consists of images of different rock art panels across the African continent. There are different image types found in the database, showing paintings, drawings, ethnographic materials, scanned historical copies of rock art, overview images of rocks containing drawings, maps saved in .jpg format showing routes to the rock art sites, rock engravings, image of sketches and tracings. This project focuses on processing the attributes and material of the associated documented data in order to extract spatial information that can be used in the process of geocoding.

The assessment of the spatial capabilities is undertaken in two phases: the extraction of data from SARADA and assessing the extracted datasets in order to identify suitable spatial information for the geocoding process.

Information about all rock art sites in Limpopo are extracted from SARADA to a JSON file using a script written in JavaScript. The script itself is provided in Appendix A. The extracted information is organised into specific fields containing province, district, how to find site (i.e. addresses, nearest town and map sheets) and other fields that have been captured as part of the metadata in the archive.

An excerpt of some extracted information of a rock art site is illustrated in the example below:

[

"alternative_site_id": null,

"mountain_region": null,
"brief_description_of_site": null,
"how_to_find_the_site": "Follow the Steillopbrug-Zwart road. Large hill. Site obvious from well out on the South side of the hill. A dark recess up a steep face near the top.",
"security_level": null,
"site_name": "Breda I 373 LR",
"site_coordinates_longtitude": 28.69722,
"district": "Mokerong 2",
"conservation_interventions": null,
"site_id": "RSA BRE1",
Type : "engraving": [],
"technique": "Painting",
"site_coordinates_latitude": -23.36389,
"when_was_site_visited_latest": null,
"nearest_town": "Monte cristo"
"date": "1999-02-04",
"images":
{
"date": "1999-02-04",
"collection_id": 2,
"has_rights": false,
"id": 20067,
"image_id": "RSA BRE1 1",
"image_large": "http://www.sarada.co.za/api/library/images/1200x800/RSA_BRE1_1",
"image_thumb": "http://www.sarada.co.za/api/library/images/300x200/RSA_BRE1_1",
"isFavourite": false,
"slug": "RSA-BRE1-1",
"url": http://www.sarada.co.za/api/images/RSA-BRE1-1 }

3.2.2 Geographical data quality assessment

The assessment which includes both the spatial and non-spatial information, consisted in verifying whether fields such as 'site_name', 'district', 'postal_codes', 'site_coordinates_longtitude', 'site_coordinates_longtitude' and the 'how_to_find_the_site', 'Nearest town' contain relevant information for each rock art site recorded in SARADA. Quite a considerable number of recorded rock art sites had missing values (null values) for the different fields of database (SARADA).Tables 3 and Table 4 below illustrate how the various pieces of information of selected rock art sites have been extracted and categorised under their respective fields in SARADA.

Table 3: Rock art sites with coordinates

Site name	District	How to find site	Latitude	Longitude	Postal codes	Nearest town
RSA BRE1 8	Mokerong 2	Null	-23.36389	28.69722	0600	Monte Christo
RSA CRD1	Ellisras	Null	-23.67083	27.93056	0555	Onverwacht
Edmondsburg I 32 MS	Soutpansberg	Null	-22.30972	29.28389	0600	GaMonyebodi
Goedgedacht I 184 KQ	Waterberg	Null	-24.08222	27.83611	0530	Vaalwater
Klipplaat I 34 KR	Waterberg	Null	-24.04861	28.31389	0530	Vaalwater
Bangor II 759 MS	Soutpansberg	Null	-22.98194	29.55028	0530	GaMonyebodi
Maria I 564 LR"	Ellisras	Null	-23.80833	27.53028	0530	GaMonyebodi
Modena I 13 MS	Soutpansberg	Null	-22.25419	29.17378	0530	GaMonyebodi
Bonne Esperance XXI 356 LR	Seshego	Null	-23.29722	28.80722	0712	GaMonyebodi
Bonne Esperance XXII 356 LR	Seshego	Null	-23.29583	28.80833	0712	GaMonyebodi
Greefswald VII 37 MS	Messina (Musina)	Null	-22.23056	29.37028	0900	GaMonyebodi
Greefswald X 37 MS	Messina (Musina)	Null	-22.21361	29.35472	0900	GaMonyebodi
Schroda V 46 MS	Messina (Musina)	Null	-22.21111	29.41389	0900	GaMonyebodi
Dwaalhoek II 185 KQ	Waterberg	Null	-24.07083	27.81889	0530	Vaalwater
Dwaalhoek III 185 KQ	Waterberg	Null	-24.07	27.81944	0530	Vaalwater
Dwaalhoek VI 185 KQ	Waterberg	Null	-24.06806	27.81944	0530	Vaalwater

Graaflust VI 637 LR	Waterberg	Null	-23.95	28.26944	0530	GaMonyebodi
Kliphoek III 636 LR	Waterberg	Null	-23.91889	28.25833	0530	GaMonyebodi
Schroda VI 46 MS	Messina (Musina)	Null	-22.21111	29.41389	0530	GaMonyebodi

Table 4: Rock art sites with postal codes

Site name	District	"How to find site" - Address	Latitude	Longitude	Postal code
Too Late XXVIII 359 LR	Bochum	null	null	null	0709
Too Late XXXIX 359 LR	Bochum	null	null	null	0709
Doppie III 95 MT	Venda (Thohoyandou)	null	null	null	0950
Alicedale II 138 MT	Messina (Musina)	null	null	null	0900
Amonda III 161 MT	Messina (Musina)	null	null	null	0900
David VI 160 MT	Messina (Musina)	null	null	null	0900
Dawn I 71 MT	Messina (Musina)	null	null	null	0900
Doreen I 108 MT	Messina (Musina)	null	null	null	0900
Schroda XVI 46 MS	Messina (Musina)	null	null	null	0900
Millstream VIII 358 LR	Seshego	null	null	null	0699
Edmondsburg II 32 MS	Soutpansberg	null	null	null	0900
Hackthorne V 30 MS	Soutpansberg	null	null	null	0900

Although there are 1180 total records of rock art sites in the Limpopo province, only rock art sites with well-defined post (ZIP) codes (and without geographic coordinate locations) were geocoded using the postal code region shapefile as a reference dataset. The attribute accuracy and completeness were assessed using a selected sample dataset and results shown in Results and Analysis.

3.2.3 Geocoding

For this research, ESRI's address locator was used to geocode against a table of standardized information obtained from the rock art database. The main method of geocoding that was used in this research is by postal code. Geocoding was performed using reference layers, these layers were obtained from the South African Post Office (SAPO), NGI (National Geospatial Information) and Google maps as a base map reference.

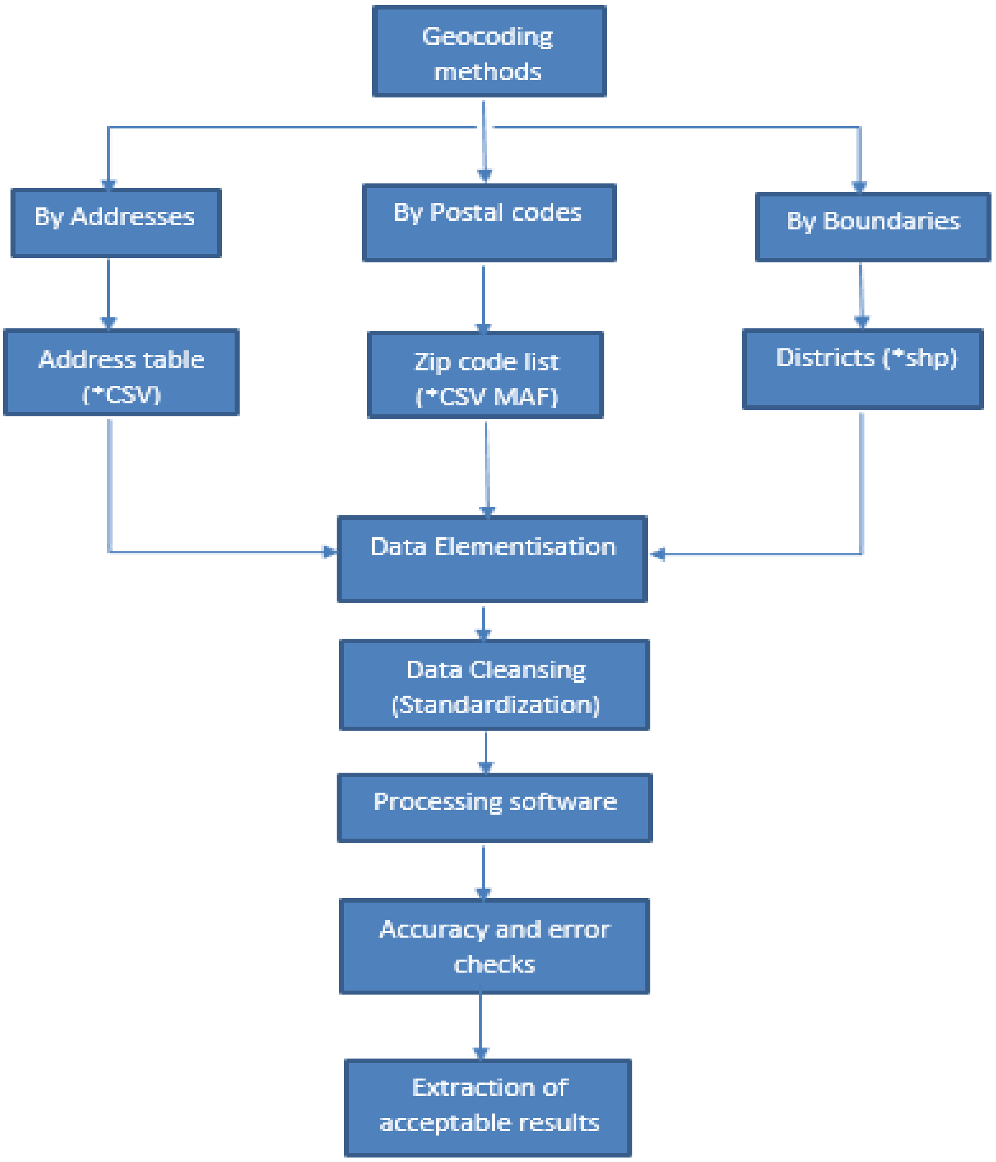


Figure 3 : Geocoding Framework

The diagram above describes a broad workflow for geocoding data. "The process is followed in order to reach an acceptable confidence level and understand the potential error rates in the data; it is broken down into steps". The first step was to separate extracted information into three different categories as mentioned above and remove information with null values, the data was extracted to a *CSV file format.

The reason for using a *CSV is because ESRI ArcGIS can read the file. The data elementisation and cleansing step involved removing characters such as punctuation symbols that will hinder processing from mentioned software. (Van Rensburg, 2015) mentions that the supported level for South Africa is level 2, which ESRI considers to be a compromising quality geocoding experience (Esri, 2013)

The following structure in figure 4 was used in the geocoding process using ESRI ArcMap Desktop, first step was to create an address locator file using datasets taken from the National Geospatial Information (NGI), Statistics South Africa (Stats_SA), South African Postal office (SAPO) and the Limpopo Survey General (Department of Rural Development and Land Reform). In South Africa there are no address locators available/accessible from public GIS data sources, an option was to purchase the file from ESRI South Africa. A geodatabase was created to add all features needed to complete the process of geocoding along with the feature classes and results are discussed in Chapter 4.

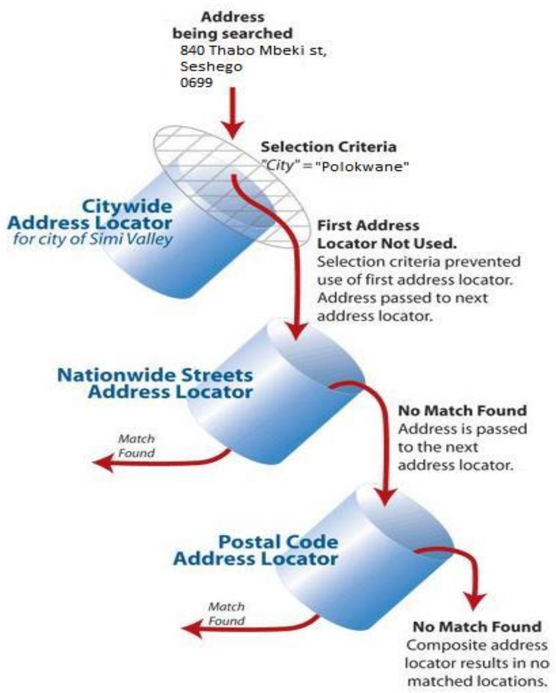


Figure 4 : A workflow diagram showing the geocoding routine (ESRI, 2013)

4. Results and Discussion

This chapter discusses the output results of data quality and geocoding rock art sites in the study area. There are a total number of 1180 of records extracted in Limpopo, 400 sites with coordinates, 780 without coordinates, 480 sites with "how to find sites" as a description for addresses and all 1180 sites contain postal codes and district boundaries (see Appendix B).

4.1 Data quality assessment

Attribute accuracy

In several literatures, there has not been a technical or quantitative method used to assess the attribute accuracy, by visually inspecting the data from the reference file and sample data, it is seen that there are many features lacking attribute information compared to the reference dataset. The district attributes in the rock art data are not named correctly as compared to reference data same goes to the nearest town. This misclassification affected overlaying the features for geocoding.

Completeness

For this study, the completeness was investigated for the rock art sites, based on the assumption that the NGI and SAPO dataset is complete. The completeness was computed by dividing the total number of towns, districts and postal codes from the reference dataset and the study area.

% Completeness = $\frac{Reference\ Datasets}{Sample\ Datasets}\ X\ 100$

% Completeness (Districts) = 5/10 x 100 = 50%

% Completeness (Towns) = 6/8 x 100 = 50%

% Completeness (Postal codes) = 1260/1180 x 100 = 106%

For both the districts and towns there is only 50% completeness of the rock art dataset compared to the reference file this is omission of data while for postal codes there is commission of data.

4.2 Geocoding results

This section explains results obtained using ESRI ArcGIS to geocode rock art sites. On the map below on figure 5, the Waterberg district was assessed in closer details, the map shows geocoded sites on the central of postal boundaries of the reference dataset, On the map shown, geographic coordinate locations are assigned to the centroid of the parcel.

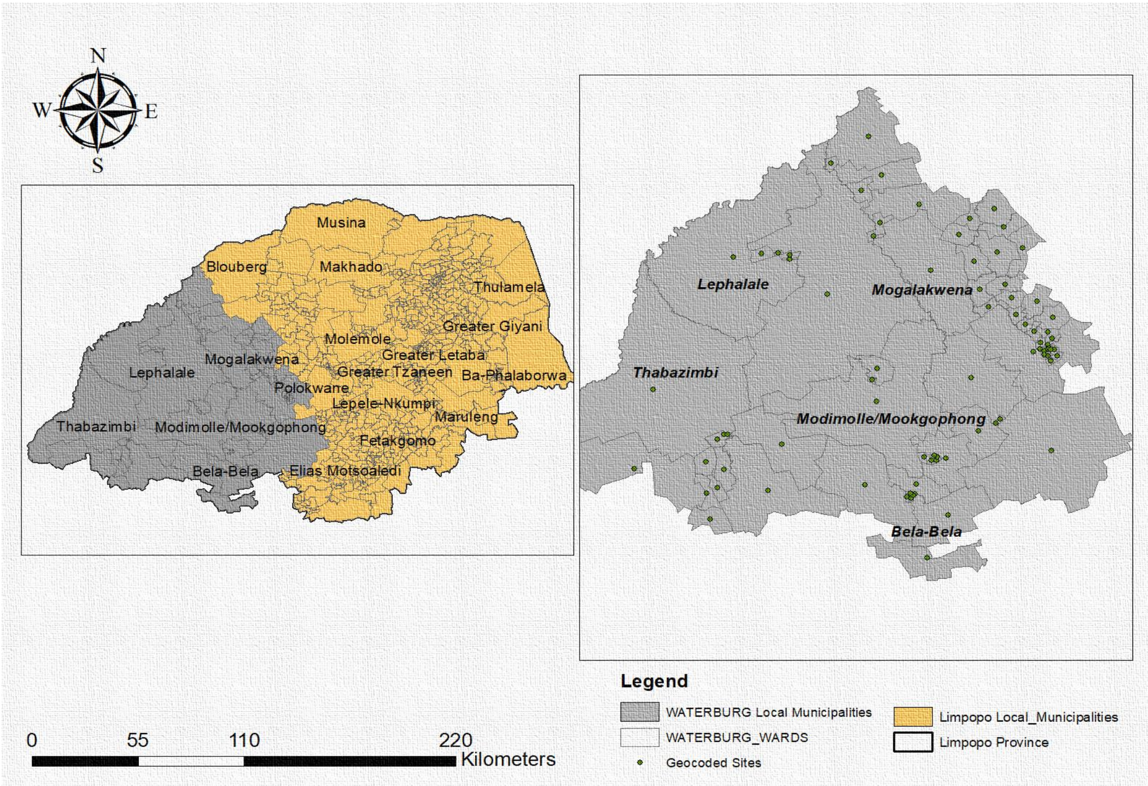


Figure 5 : Mapped rock art sites

Using the three standard geocoding methods proved to be feasible to assign rock art sites (x, y) coordinates but at a postal code level, the variability in match rates also points to the influence of the data input quality. This strongly outlines that the quality of the input data and the quality of the geocoding process are both important contributors to the final output. Geocoding using addresses is not possible on rock art sides as there is no address data. Geocoding using postal/zip codes is feasible but with poor accuracy as this uses centroid points of postal code boundaries. Geocoding using district boundaries or administrative is the least accurate as mentioned in literature and cannot be used for navigational purposes.

One of the main problems identified with the postal code files was that many postal codes, as well as many place names, are mapped to multiple coordinate points, as shown in figure 5. The reason being that town/suburbs/cites can encompass more than one postal code. This means that a greater city region such as Polokwane, will result in dozens of related postal codes and geographic points. This is further shown below when using the Geocoding API on Google Maps.

The results obtained correlate with existing literature on geocoding with postal code i.e. rural areas have more inaccuracies in terms of geocoding outcomes when compared to urban areas (Healy and Gilliland, 2012, Ng et al., 1993, Zandbergen, 2008).

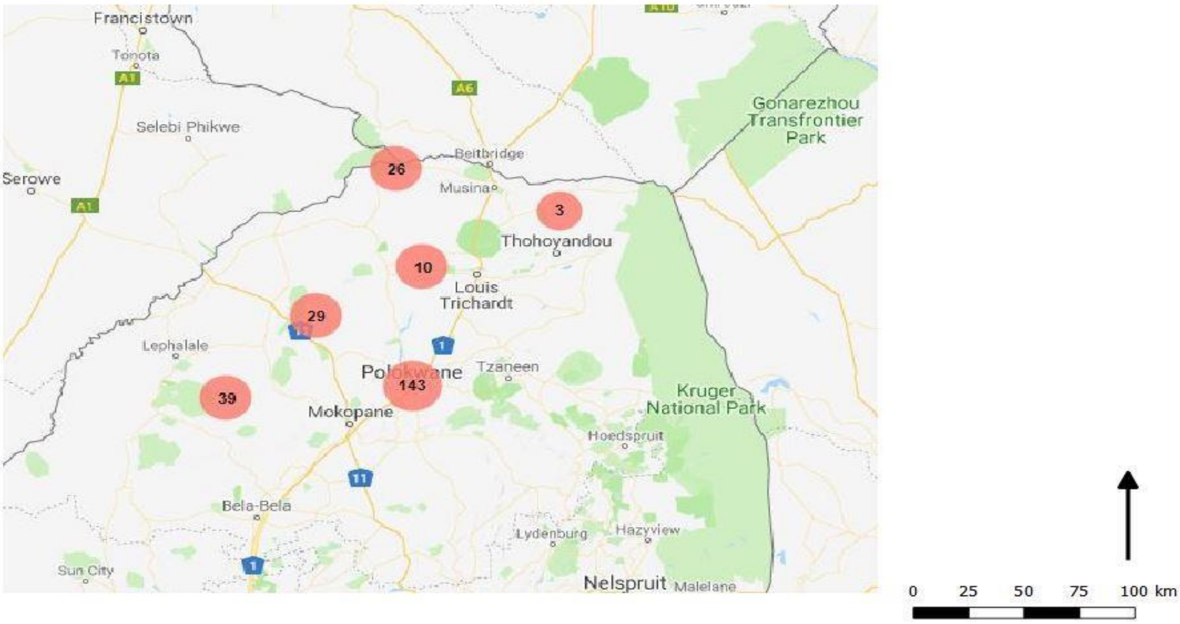
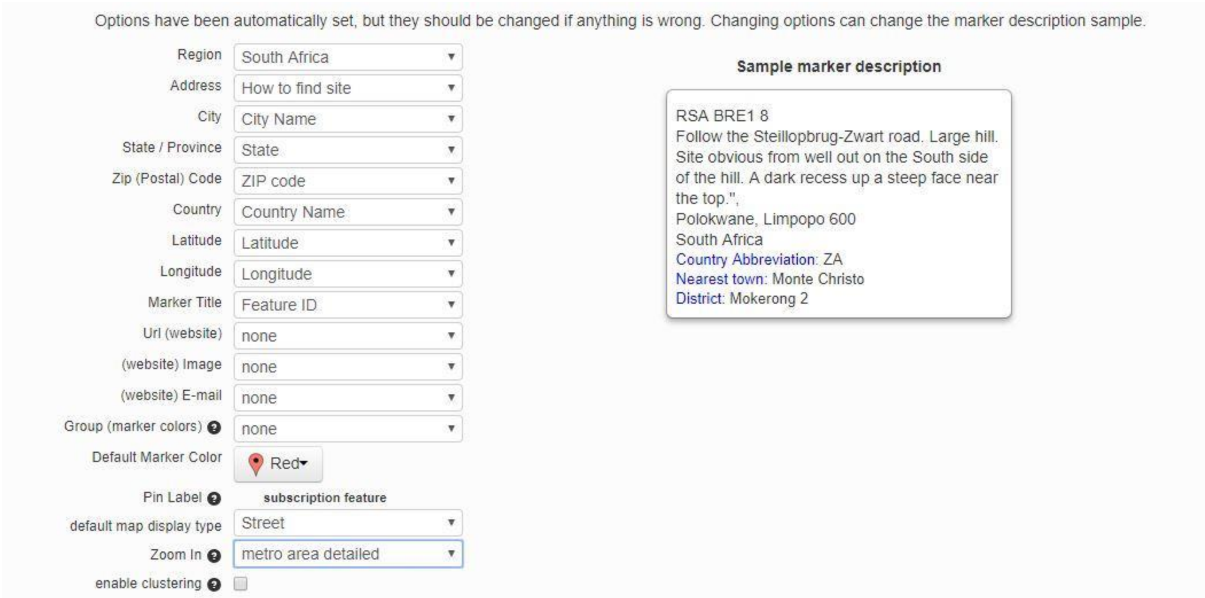


Figure 6: Geocoding Rock Art Sites using Google Maps

Google Maps service API was used as secondary check, this was to compare with the results from ESRI ArcGIS, a total of 250 sites were geocoded as shown in the figure. But there were error returns in the first run. Only 26 sites in the northern area of Mapungubwe produced a complete match rate, in the Polokwane area showing 143 sites, it showed to have been misallocated. With this geocoding method, it firstly looks at address and if not available it passes to postal codes and lastly it uses the place name or town to place the points.



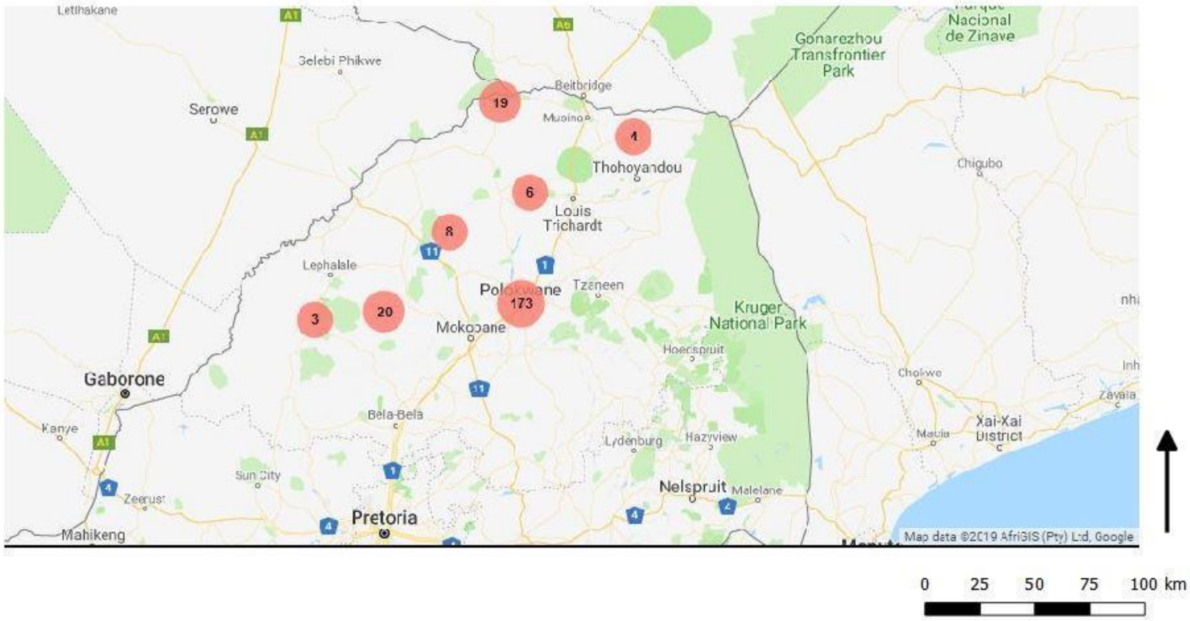


Figure 7: Geocoding Rock Art Sites using Google Maps

For the second run of using this geocoding method as a reference check, only 480 out of 1180 were geocoded. There were a lot of problems discovered using this service, for example Google uses place names or nearest town as a field to provide geocoding results if there is no address data. Hence a total number of 316 sites fall under "Polokwane" city from both figure 6 and figure 7. Rock art sites are mostly found in mountainous or rocky areas and it is almost impossible to find these sites in a city.

Match rate = 730 (Geocoded)/1180 (Total) x 100

Match rate = 62% (Based on the geocoding scale, this was measured on a road level)

A related study was conducted by Duncan et al. (2011) that is based on online geocoding services (Google and MapQuest) to validate the geocoding results from those obtained by a commercial GIS (ArcGIS). He concluded that Google has a higher match rate and its geocoding results are likely to be closer to the ones obtained from ArcGIS but in the case of geocoding using postal codes, Google was found to not be reliable in this study due to the nature of the data.

5. Conclusion and Recommendations

The aim of this study was to assess the suitability of textual information extracted from SARADA for the purpose of geocoding rock art sites in Limpopo. SARADA is a very large database that encompasses African rock art history. Limpopo was selected as the study area in order to cover a variety of rock art forms and the rich cultural history. A JavaScript was used to extract essential information from the database. The spatial content of the extracted data was assessed, the data was categorised using summary tables with respectable allocations i.e. site names; district; "How to find site ", site coordinates, postal codes and nearest town.

The data quality is a huge contributor to the geocoding rate, as per literature all geographical data quality assessment was taken into consideration but only two data quality assessments were performed due to the nature of data. This paper extends the current state of knowledge on geographical data quality assessment by focusing on completeness and attribute accuracy. A manual record matching procedure was used to analyse the extend of the data quality and the percentage of matches was demonstrated in terms of completeness of the data.

After assessing the data quality, a geocoding procedure was performed using ESRI ArcGIS software and Google Earth for secondary check. Geocoding of a rural space was problematic because of the rural routes which cannot be geocoded. Rural routes and postal boxes are assigned the coordinates of the zip code centroids and results in misclassification. Even when rural addresses can be geocoded, results of our study and others indicate that positional errors are likely to be larger than for addresses in towns. Furthermore, as indicated by this study the choice of a geocoding method affects both geocoding success and the positional error. Therefore, careful consideration should be given to the approach used for geocoding.

In conclusion, the information within the South African Rock Art Digital Archive can be geocoded only at the postal code (or municipal and district levels) but with poor quality and accuracy. To improve the chances of locating the rock art sites, the attributes of "How to find site" can be used in addition in order to assist the RARI community/interested individuals to further locate sites. Due to the poor level accuracy obtained in the geocoding routine, this study recommends that the solution in improving the positional error lies upon capturing the correct coordinates of rock art sites in the field.

Assessing the spatial capabilities of the South African Rock Art Archive (SARADA) in purpose geocoding the rock art sites will lead to positive contributions to the rock art research.

6. References

Bonner, M. R., Han, D., Nie, J., Rogerson, P., Vena, J. E., & Freudenheim, J. L. (2003). Positional accuracy of geocoded addresses in epidemiologic research. *Epidemiology, 14(4), 408–412.*

Campbell, C., 2013. *Art in crisis: contact period rock art in the south-eastern mountains of southern Africa*. PhD thesis, University of the Witwatersrand.

Cayo, M. R., & Talbot, T. O. (2003). Positional error in automated geocoding of residential addresses. *International Journal of Health Geographics, 2(10).*

Charif, O., Omrani, H., Klein, O., Schneider, M. & Trigano, P.,2010. A method and a tool for geocoding and record linkage. Geoscience and Remote Sensing (IITA-GRS), Second IITA International Conference on, 2010. IEEE, 356-359.

Chippindale, C. and Nash, G., 2004. Pictures in place: approaches to the figured landscapes of rock-art. *The figured landscapes of rock-art: looking at pictures in place,* pp.1-36.

Coetzee, S. & Cooper, A. K. 2007. What is an address in South Africa? *South African Journal of Science,* 103**,** 449-458.

Collins, S. E., R. P. Haining, I. R. Bowns, D. J. Crofts, T. S. Williams, A. S. Rigby, and D. M. Hall. 1998. Errors in postcodeto enumeration district mapping and theireffect on small area analyses of health data. Journal of Public Health Medicine 20(3): 325-30.

Conolly, J., & Lake, M. (2006). Geographical Information Systems in Archaeology. Cambridge: Cambridge University Press.

Cura, R., Dumenieu, B., Abadie, N., Costes, B., Perret, J. & Gribaudi, M. 2018. Historical collaborative geocoding. ISPRS International Journal of Geo-Information, 7, 262. Available at https://www.mdpi.com/2220-9964/7/7/262 [17 May 2017]

Dao, J.T., 2015. *A Comparison of Address Point and Street Geocoding Techniques in a Computer Aided Dispatch Environment* (Master's Dissertation University of Southern California).

Davis Jr., C. A., F. T. Fonseca, and K. A. De Vasconcelos Borges. 2003. A flexible addressing system for approximate geocoding, GeoInfo 2003: Proceedings of the Fifth Brazilian Symposium on GeoInformatics, Campos do Jordão, São Paulo, Brazil, October 2003.

Dearwent, S. M., Jacobs, R. J., & Halbert, J. B. (2001). Locational uncertainty in georeferencing public health datasets. *Journal of Exposure Analysis and Environmental Epidemiology, 11, 329–334.*

Docan, D. C. 2013. Spatial data quality assessment in GIS, ed. A.C Badea, F.R Ribeiro, C. Grea, I. Veres. 08-10 October 2013, *In Proceedings of the 1st European Conference of Geodesy and Geomatics Engineering* pp. 105–112, Antalya, Turkey WSEAS Press

Duncan, D. T., M. C. Castro, J. C. Blossom, G. G. Bennett, and S. L. Gortmaker. 2011. *Evaluation of the positional difference between two common geocoding methods.*

Dramowicz, E. (2004). *Three standard geocoding methods.* Directions Magazine., Available at: http://www.directionsmag.com/article.php?article_id=670&trv=1 [22 October 2008]

Eastwood, E. B., and Smith, B. W. (2005). Fingerprints of the Khoekhoen: geometric and handprinted rock art in the Central Limpopo Basin, South Africa. *South African Archaeological Goodwin Series* **9:** 63–76.

ESRI, G.I.S., 2014. Dictionary. Definitions for GIS terms related to operations such as analysis, GIS modeling and web-based GIS, cartography, and Esri software. Available at https://support.esri.com/en/other-resources/gis-dictionary [20 August 2017]

ESRI. 2015. Geocode Services. Available at http://www.science.smith.edu/sal/wp-content/uploads/sites/13/2015/03/Geocoding-Guide.pdf [29 July 2018]

El-Hakim, S. F., Fryer, J. G. and Picard, M. 2004, Modelling and visualization of aboriginal rock art in the Baiame cave. *ISPRS International Archives of Photogrammetry and Remote Sensing*, 35(5): 990-995.

Gaiz, Kk., 2013. *Implementation of a geodatabase and GNSS data collection workflow for the archaeology program at the flagstaff area national monuments*. Msc thesis, Northern Arizona University. Available at: https://s3.amazonaws.com/academia.edu.documents/33120645/GaizPracticumFinal_20131028.pdf Accessed [29 July 2017]

Gatrell, A. C. 1989. On the spatial representation and accuracy of address-based data in the United Kingdom. Int. Journal of Geographical Information Systems 3(4): 335-48.

Gupta, K. (updated and expanded by Sleezer, C. M. & Russ-Eft, D. F.). (2007). A practical guide to needs assessment (2nd ed.). San Francisco, CA: Pfeiffer.

Gregory, I.N. and Healey, R.G., 2007. Historical GIS: structuring, mapping and analysing geographies of the past. *Progress in human geography*, *31*(5), pp.638-653.

Grossmann, C. H. 2005. *A web-based GIS as a management tool for rock art sites in the uKhahlamba-Drakensberg Park*., M.Sc. thesis, University of Kwazulu-Natal. Available at http://ukzn-dspace.ukzn.ac.za/handle/10413/3409 [12 April 2017]

Hall, N., Koeneman, T., Sullivan, S. and Taçon, P.S., 2016. Rock art: a cultural treasure at risk. *Rock Art Research*, *33*(1), 116-117.

Haubt, R.A., 2015. The global rock art database: developing a rock art reference model for the RADB system using the CIDOC CRM and Australian heritage examples. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* 1, 89–96

Harries, Keith. 1999. Mapping Crime: Principle and Practice. Available at https://www.ncjrs.gov/html/nij/mapping/index.html [14 August 2018]

Hardtke, F. E. 2017. Rock art interpretive approaches–Devising frameworks to maximally utilise independent lines of evidence. *Journal of Arid Environments*, 143, 22-27.

Healy MA, Gilliland JA. Quantifying the magnitude of environmental exposure misclassification when using imprecise address proxies in public health research. *Spatial and Spatio-temporal Epidemiology* 2012; 3: 55–67.

Hurley, S.E., Saunders, T.M., Nivas, R., Hertz, A. and Reynolds, P., 2003. Post office box addresses: a challenge for geographic information system-based studies. *Epidemiology*, *14*(4), pp.386-391.

Kacyra, B., 2009. CyArk 500–3D documentation of 500 important cultural heritage sites. In *Photogrammetric Week* (Vol. 9, pp. 315-320).

Karimi, H. A., & Durcik, M. (2004). *Evaluation of uncertainties associated with geocoding techniques.* Computer-Aided Civil and Infrastructure Engineering, 19, 170–185.

Kellison MT (2012) *Address Points and a Master Address File: Improving Efficiency in the City of Chino*. MSc thesis, Unversity of Southern California, CA, USA. Available at https://spatial.usc.edu/wp-content/uploads/2014/04/KellisonMichaelThesis.pdf

Khan A, Vasardani M, and Winter S 2013 Extracting spatial information from place descriptions. In *Proceedings of the First ACM SIGSPATIAL International Workshop on Computational Models of Place*, Orlando, Florida

Kollen, C., Dietz, C., Suh, J. and Lee, A., 2013. Geospatial data catalogs: Approaches by academic libraries. *Journal of Map & Geography Libraries*, *9* (3), pp.276-295.

Krieger, N., J. T. Chen, P. D. Waterman, M.-J. Soobader, S. V. Subramanian, and R. Carson. 2002a. Geocoding and monitoring of U.S. socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter? American Journal of Epidemiology 156(5): 471-82.

Lee, H. D., & Nelson, O. W. (2006). Instructional analysis and course development. Homewood, IL: American Technical Publishers, INC.

Leathwick, J. (2000, November). *Predictive Models of Archaeological Site Distributions in New Zealand.* Wellington: New Zealand Department of Conservation. Department of Conservation: Available at http://www.doc.govt.nz/documents/science-and-technical/ir181.pdf

Levine, N. and Kim, K.E., 1998. The location of motor vehicle crashes in Honolulu: a methodology for geocoding intersections. *Computers, environment and urban systems*, *22*(6), pp.557-576.

Malizia, N. 2013. The effect of data inaccuracy on tests of space-time interaction. *Transactions in GIS*, 17**,** 426-451.

May, S.K., P.S.C. Taçon, D. Wright, M. Marshall, J. Goldhahn & I. Domingo Sanz, 2017. *The rock art of Madjedbebe (Malakunanja II)*, in B. David, P.S.C. Taçon, J.J. Delannoy & J.M. Geneste (eds) *The Archaeology of Rock Art in Western Arnhem Land, Australia* Canberra: ANU Press, 87–108.

Merryman, J.H., 1989. The public interest in cultural property. *Calif. L. Rev.*, *77*, p.339.

Messner, A. (2009). Needs Assessment and Analysis Methods. Masters thesis. University of Wisconsin-Stout. Available at http://www2.uwstout.edu/content/lib/thesis/2009/2009messnera.pdf

Myers, D. and Dalgity, A., 2012. The Middle Eastern Geodatabase for Antiquities (MEGA): an open source GIS-based heritage site inventor y and management system. *Change Over Time*, *2*(1), pp.32-57.

Ng E, Wilkins R, Perras A. How far is it to the nearest hospital? Calculating distances using the Statistics Canada Postal Code Conversion File. *Health Reports* 1993; 5(2): 179–88.

Plets, G., Verhoeven, G., Cheremisin, D., Plets, R., Bourgeois, J., Stichelbaut, B., Gheyle, W. and De Reu, J., 2012. The deteriorating preservation of the Altai rock art: assessing three-dimensional image-based modelling in rock art research and management. *Rock Art Research: The Journal of the Australian Rock Art Research Association (AURA)*, *29*(2), pp.139-145.

Ratcliffe, J. H. (2004). Geocoding crime and a first estimate of a minimum acceptable hit rate. *International Journal of Geographical Information Science, 18(1), 61–72.*

Ratcliffe, J. H. (2001). On the accuracy of TIGER-type geocoded address data in relation to cadastral and census areal units. *International Journal of Geographical Information Science, 15(5), 473–485.*

Schootman, M., Sterling, D. A., Struthers, J., Yan, Y., Laboube, T., Emo, B., & Higgs, G. (2007). Positional accuracy and geographic bias of four methods of geocoding in epidemiologic research. *Annals of Epidemiology, 17(6), 464–470.*

Sheehan, T. J., S. T. Gershman, L. MacDougal, R. A. Danley, M. Mroszczyk, A. M. Sorensen, and M. Kulldorff. 2000. Geographic surveillance of breast cancer screening by tracts, towns and zip codes. Journal of Public Health Management Practices 6: 48-57.

Strickland, M. J., Siffel, C., Gardner, B. R., Berzen, A. K., & Correa, A. (2007). Quantifying geocode location error using GIS methods. *Environmental Health, 6(10).*

Swift, J., Goldberg, D. & Wilson, J. 2008. Geocoding best practices: review of eight commonly used geocoding systems. *Los Angeles, CA: University of Southern California GIS Research Laboratory*.

Tian, Q., Ren, F., Hu, T., Liu, J., Li, R. & Du, Q. 2016. Using an optimized chinese address matching method to develop a geocoding service: A case study of shenzhen, china. *ISPRS International Journal of Geo-Information*, 5**,** 65.

Titus, J.M., 2016. *Peoples of Washington Historical Geographic Information System: Geocoding Culture using Archival Standards* (PhD thesis), University of Southern California). Available at https://spatial.usc.edu/wp-content/uploads/2016/04/Titus-Jennifer.pdf

Veregin, H., 1999. Data quality parameters. Geographic Information Systems: Principles and Technical Issues. MF Goochild, PA Longley, DJ Maguire, and DW Rhind. Second (ed.) Vol. 1.

Vodeb, V. & Zakrajsek, F. 2013. Geocoded Digital Cultural Content. Roma: Linked Heritage project.

Ward, M. H., Nuckols, J. R., Giglierano, J., Bonner, M. R., Wolter, C., Airola, M., et al. (2005). Positional accuracy of two methods of geocoding. *Epidemiology, 16(4), 542–547.*

Wagtendonk, A. V. (2009). Past in Place: The Role of Geo-ICT in Present-day Archaeology. Geospatial technology and the role of location in science, 59-86.

Wienhold, M. L., & Robinson, D. W. (2017). GIS in rock art studies. In B. David & I. J. McNiven (Eds.), *The Oxford handbook of the archaeology and anthropology of rock art* (pp. 1–30). Oxford: Oxford University Press. Available at https://doi.org/10.1093/oxfordhb/9780190607357.013.1. [12 May 2018]

Wikin, B., Altschuld, J. (1995). *Planning and conducting needs assessments: A practical guide*. Thousands Oaks, CA: Sage Publications. https://ctb.ku.edu/en/table-of-contents/assessment/assessing-community-needs-and- resources/develop-a-plan/main Accessed [05 February 2019]

Wheatley, D., & Gillings, M. (2002). *Spatial Technology and Archaeology: The Archaeological Applications of GIS*. London: Taylor & Francis.

Whitsel, E. A., Rose, K. M., Wood, J. L., Henley, A. C., Liao, D., et al. (2004). Accuracy and repeatability of commercial geocoding. *American Journal of Epidemiology, 160(10), 1023–1029.*

Whitsel, E. A., Quibrera, P. M., Smith, R. L., Catellier, D. J., Liao, D., Henley, A. C., & Heiss, G. (2006). Accuracy of commercial geocoding: *Assessment and implications. Epidemiological Perspectives and Innovations, 3(8).*

UNESCO (2008): Intergovernmental Committee for the Protection of the World Cultural and Natural Heritage: *Operational Guidelines for the Implementation of the World Heritage Convention*. Available at http://whc.unesco.org/archive/opguide96.pdf

Zandbergen, P.A. 2008. A comparison of address point, parcel and street geocoding techniques. *Computers, Environment and Urban Systems*, 32, 214-232.

Zandbergen, P.A., 2008. Positional accuracy of spatial data: Non-normal distributions and a critique of the national standard for spatial data accuracy. *Transactions in GIS*, 12(1), pp.103-130.

Zandbergen, P. A. (2007). Influence of geocoding quality on environmental exposure assessment of children living near high traffic roads. *BMC Public Health, 7(37).*

Zhan, F. B., Brender, J. D., De Lima, I., Suarez, L., & Langlois, P. H. (2006). Match rate and positional accuracy of two geocoding methods for epidemiologic research. *Annals of Epidemiology, 16(11), 842–849.*

Zimmerman, D. L., Fang, X., Mazumdar, S., & Rushton, G. (2007). Modeling the probability distribution of positional errors incurred by residential address geocoding. *International Journal of Health Geographics, 6(1).*

Zimmerman, D. L. 2008. Estimating the intensity of a spatial point process from locations coarsened by incomplete geocoding. *Biometrics* 64 (1):262-270.

Zimmerman, D. L., and J. Li. 2010. The effects of local street network characteristics on the positional accuracy of automated geocoding for geographic health studies. *International journal of health geographics* 9 (1):10.

Zimmerman, D. L., J. Li, and X. Fang. 2010. Spatial autocorrelation among automated geocoding errors and its effects on testing for disease clustering. *Statistics in medicine* 29 (9):1025-1036.

Websites

Limpopo Information Directory. 2018:

https://www.southafricainfo.co.za/country/province/10/limpopo Accessed 12 March 2018

Geography and History

http://www.southafricanewyork.net/consulate/PocketGuide2017/provinces1617.pdf

[Accessed 12 March 2018]

World Heritage Sites. 2018: https://www.aluka.org/heritage/partner/XSTRARI

[Accessed 10 February 2018]

The Rock Art Research Institute. 2018: https://artsandculture.google.com/partner/the-rock-artresearch-institute-university-of-the-witwatersrand-johannesburg
[Accessed 18 September 2018]
Spatially enabling the data: What is geocoding?
https://www.ncjrs.gov/html/nij/mapping/ch4_3.html.[Accessed 12 July 2018]
Geocoding Services .2018.
http://resources.esri.com/help/9.3/arcgisserver/apis/soap/SOAP_Geocode_Overview.htm.GA
[Accessed 19 September 2018]
Google Maps Platform
https://developers.google.com/maps/terms#section_10_12 [Accessed 21 September 2018]
The Mirrar rock art Project: https://www.researchgate.net/project/Mirarr-Gunwarddebim-theMirarr-Rock-Art-Project [Accessed 05-November 2018]
Uncertainty in Historical GIS. 2017
https://www.geocomputation.org/search?q=uncertainty in historical gis/mamani-sanchezbertolotto [Accessed 10 January 2019]
Gis User Survey and Needs Analysis Study Final Report
http://www.dbw.ca.gov/csmw/pdf/GISUserSurvey_FinalReport07182011.pdf [Accessed 03 March 2019]

7. Appendices:

Appendix A : JAVASCRIPT used to extract data from SARADA database

```
“import java.io.FileNotFoundException”;
“import java.io.IOException”;
“import java.io.PrintWriter”;
“import org.json.JSONArray”;
“import org.json.JSONObject”;
“import okhttp3.OkHttpClient”;
“import okhttp3.Request”;
“import okhttp3.Response”;
“public class” SaradaExtract {
“public static void main(String[] args)” {
SaradaExtract extract = new SaradaExtract();
String searchUrl =
"http://www.sarada.co.za/api/sites?search=%7B%22search%22:%22Limpopo%22%7D";
String outputFilepath = "C:/tmp/";
extract.getSitesDetails(searchUrl, outputFilepath);
}
public void getSitesDetails(String searchUrl, String outputFilepath){
String[] siteUrlList = getSiteUrlList(searchUrl);
JSONArray siteDetailsList = getSiteDetails(siteUrlList);
String[] siteSpatailDataList = extractSpatialData(siteDetailsList);
writeJSONArrayTofile(siteDetailsList,outputFilepath);
writeArrayDataToFile(siteSpatailDataList, outputFilepath);
}
public String[] getSiteUrlList(String searchUrl){
  String sResponseBody
  JSONObject jsonResponseBody
  int totalNumberOfSite
  String[] saSiteUrsList
  = makeRequest(searchUrl);
  = new JSONObject(sResponseBody);
  = jsonResponseBody.getInt("total");
  = new String[totalNumberOfSite];
  int numOfRequiredSubsequetRquest = jsonResponseBody.getInt("last_page");
  String nextReguestUrl
  jsonResponseBody.getString("next_page_url");
  JSONArray jaResonpseSites
  String[] saResponseSites
  int sitesAdded
  =
  = jsonResponseBody.getJSONArray("data");
  = extractSiteUrlData(jaResonpseSites);
  = 0;
for (String siteUrl : saResponseSites){
saSiteUrsList[sitesAdded] = siteUrl;
53
sitesAdded++;
break;
}
```

```
for (int nextRequestCounter = 2; nextRequestCounter <=
numOfRequiredSubsequetRquest; nextRequestCounter++) {
  sResponseBody
  jsonResponseBody
  = makeRequest(nextReguestUrl);
  = new JSONObject(sResponseBody);

  jaResonpseSites
  saResponseSites
  = jsonResponseBody.getJSONArray("data");
  = extractSiteUrlData(jaResonpseSites);
for (String siteUrl : saResponseSites){
saSiteUrsList[sitesAdded] = siteUrl;
sitesAdded++;
}
if (nextRequestCounter < numOfRequiredSubsequetRquest) {
nextReguestUrl = jsonResponseBody.getString("next_page_url");
}
}
System.out.println("Retrived Site Urls");
return saSiteUrsList;
}
public String[] extractSiteUrlData(JSONArray sites){
int numOfSites = sites.length();
String[] sitesUrl = new String[numOfSites];
for (int index = 0; index < sites.length(); index++) {
sitesUrl[index] = sites.getJSONObject(index).getString("url");
}
return sitesUrl;
}
public JSONArray getSiteDetails(String[] siteUrlList) {
JSONArray jaSiteDetailsList = new JSONArray();
String sResponseBody;
JSONObject jsonResponseBody;
for (String siteUrl : siteUrlList) {
sResponseBody = makeRequest(siteUrl);
jsonResponseBody = new JSONObject(sResponseBody);
jaSiteDetailsList.put(jsonResponseBody);
```

```
}
System.out.println("Retrived Site Details");
return jaSiteDetailsList;
}
public String[] extractSpatialData(JSONArray siteDetailsList){
int listSize = siteDetailsList.length();
54
String[] saSpatialData = new String[listSize];
JSONObject jSite;
String siteRecord;
String site_id;
String site_name;
String site_coordinates_latitude;
String site_coordinates_longtitude;
String country;
String province;
String district;
String nearest_town;
String how_to_find_the_site;
String map_sheet;
for “(int index = 0; index < listSize; index++)” {
“jSite = siteDetailsList.getJSONObject(index);
site_id = jSite.getString("site_id");
site_name = jSite.getString("site_name");
site_coordinates_latitude = jSite.get("site_coordinates_latitude").toString();
site_coordinates_longtitude =
jSite.get("site_coordinates_longtitude").toString();
country = jSite.get("country").toString();
province = jSite.get("province").toString();
district = jSite.get("district").toString();
nearest_town = jSite.get("nearest_town").toString();
map_sheet = jSite.getJSONArray("map_sheet").toString();
how_to_find_the_site =
jSite.get("how_to_find_the_site").toString().replaceAll("\n", " ");”
```

```
siteRecord = site_id
+ "|" + site_name
+ "|" + site_coordinates_latitude
+ "|" + site_coordinates_longtitude
+ "|" + country
+ "|" + province
+ "|" + district
+ "|" + nearest_town
+ "|" + map_sheet
+ "|" + how_to_find_the_site;
saSpatialData[index] = siteRecord;
}
return saSpatialData;
}
public void writeJSONArrayTofile(JSONArray jsonArray, String Filepath){
try {
PrintWriter out = new PrintWriter(Filepath + "JSONSitesDetails.json");
out.print(jsonArray.toString());
55
out.”close”();
} catch (FileNotFoundException e) {
// TODO Auto-generated catch block
e.printStackTrace();
}
} “
public void writeArrayDataToFile(String[] list, String Filepath){
PrintWriter out”;
try {
“out = new PrintWriter(Filepath + "SitesDetails.txt")”;
“String siteRecordHeader”
="site_id|site_name|site_coordinates_latitude|site_coordinates_longtitude|country"
+
"|province|district|nearest_town|map_sheet|how_to_find_the_site";
out.println(siteRecordHeader);
```

```
for (String line : list){
out.println(line);
}
out.”close”();
} “catch (FileNotFoundException e)” {
// “TODO Auto-generated catch block
e.printStackTrace”();
}
} “
public String makeRequest”(String url){
//”final MediaType JSON = MediaType.parse”("application/json; charset=utf-8");
“OkHttpClient client = new OkHttpClient()”;
String results = "";
“Request request = new Request.Builder()”
“.url(url)”
“.get()”
“ .addHeader("content-type", "application/json")”
“ .build();”
“try” {
“Response response = client.newCall(request).execute”();
results =”response.body().string()”;
} “catch (IOException e)” {
// TODO Auto-generated “catch” block
e.printStackTrace();
} “
return results”;
}
}
```

Appendix B: Extracted sample data into standard geocoding format

Site name	Countr y	Provinc e	City	Nearest town	ZIP code	District	Latitude	Longitude	"How to find site"
RSA BRE1 8	ZA	Limpopo	Polokwane	Monte Christo	0600	Mokerong 2	-23.36389	28.69722	Follow the Steillopbrug Zwart road. Large hill. Site obvious from well out on the South side of the hill. A dark recess up a steep face near the top.",
RSA CRD1	ZA	Limpopo	Polokwane	Onverwacht	0555	Ellisras	-23.67083	27.93056	Direction from Ellisras. Take the road to Marken from Ellisras. Approximately 20km from Ellisras there is sign on the right for Zingela. Drive straight following the phone lines until you come to a sign. Follow directions to spider (manager). This road ends up at the Stubb's

									house. To get to the site follow road for about 500m where there is a track on the left. Follow the track to towards the mountain (Tafelkoppe). Walk up the old road to the Dam. Near the top ridge turn left. The shelter is along the ridge on the right.",
Edmondsburg I 32 MS	ZA	Limpopo	Polokwane	Onverwacht	0555	Soutpansberg	-22.30972	29.28389	null

Gallashiels VIII 316 LR	ZA	Limpopo	Polokwane	Blouberg	0790	Bochum	null	null	null
Goedgedacht I 184 KQ	ZA	Limpopo	Polokwane	Vaalwater	0530	Waterberg	-24.08222	27.83611	null
Klipplaat I 34 KR	ZA	Limpopo	Polokwane	Vaalwater	0530	Waterberg	-24.04861	28.31389	null
Bangor II 759 MS	ZA	Limpopo	Polokwane	Onverwacht	0530	Soutpansberg	-22.98194	29.55028	null
Maria I 564 LR"	ZA	Limpopo	Polokwane	Onverwacht	0555	Ellisras	-23.80833	27.53028	null
Modena I 13 MS	ZA	Limpopo	Polokwane	Onverwacht	0555	Soutpansberg	-22.25419	29.17378	null
Vincent I	ZA	Limpopo	Polokwane	Onverwacht	0555	Soutpansberg	null	null	null
Langbryde XIII	ZA	Limpopo	Polokwane	Bochum and Gamonyebodi	0709	Bochum	null	null	null
Millbank II 325 LR	ZA	Limpopo	Polokwane	Bochum and Gamonyebodi	0709	Bochum	null	null	null
Mont Blanc II	ZA	Limpopo	Polokwane	Driekoppies and Gamonyebodi	0709	Bochum	null	null	null
Mont Blanc III	ZA	Limpopo	Polokwane	Driekoppies and Gamonyebodi	0709	Bochum	null	null	null
Mont Blanc XV	ZA	Limpopo	Polokwane	Driekoppies and Gamonyebodi	0709	Bochum	null	null	null
Too Late III	ZA	Limpopo	Polokwane	Driekoppies and Gamonyebodi	0709	Bochum	null	null	null

Too Late XIX	ZA	Limpopo	Polokwane	Driekoppies and Gamonyebodi	0709	Bochum	null	null	Ed Eastwood
Skirbeek II 73 MT	ZA	Limpopo	Polokwane	Messina and Muswodi	0900	Messina (Musina)	null	null	null
Bonne Esperance VI 356 LR	ZA	Limpopo	Polokwane	GaMonyebod i	0712	Seshego	null	null	null
Bonne Esperance XXI 356 LR	ZA	Limpopo	Polokwane	GaMonyebod i	0712	Seshego	-23.29722	28.80722	null
Bonne Esperance XXII 356 LR	ZA	Limpopo	Polokwane	GaMonyebod i	0712	Seshego	-23.29583	28.80833	null
Greefswald VII 37 MS	ZA	Limpopo	Polokwane	null	0900	Messina (Musina)	-22.23056	29.37028	Enter Greefswald from Denstaat road. Turn right, drive along the fence until you see an old communications building (with transmitters), park here. The site is about 200m North of the building in the first valley you encounter.

Greefswald X 37 MS	ZA	Limpopo	Polokwane	null	0900	Messina (Musina)	-22.21361	29.35472	null
Greefswald XIII 37 MS	ZA	Limpopo	Polokwane	null	0900	Messina (Musina)	null	null	null
Schroda V 46 MS	ZA	Limpopo	Polokwane	null	0900	Messina (Musina)	-22.21111	29.41389	null
Dwaalhoek II 185 KQ	ZA	Limpopo	Polokwane	Vaalwater	0530	Waterberg	-24.07083	27.81889	See directions to Dwaalhoek I. Dwaalhoek II is a little further up stream on the left (West bank) of the river.

Dwaalhoek III 185 KQ	ZA	Limpopo	Polokwane	Vaalwater	0530	Waterberg	-24.07	27.81944	About 15m from Dwaalhoek II - to the right as you face the rock. For directions see Dwaalhoek I.
Dwaalhoek VI 185 KQ	ZA	Limpopo	Polokwane	Vaalwater	0530	Waterberg	-24.06806	27.81944	See Dwaalhoek I for directions to the valley. Site is approximately 800m up valley in a large overhang up stream on the west bank of the river.",

Graaflust VI 637 LR	ZA	Limpopo	Polokwane	Onverwacht	0530	Waterberg	-23.95	28.26944	3Km before reaching Graaflust II-IV on the dirt road turn left up a new, steep dirt road. Follow this for about 8km, branching right half way. This new road reaches the Lapalala fence and doubles back on itself down in to the valley. Cross the river, walk to the Lady Grey fence. Climb the steep cliff. \nLarge site just below a huge cave, along the cliff wall with a straight overhang.",
Kliphoek III 636 LR	ZA	Limpopo	Polokwane	GaMosidi	0530	Waterberg	-23.91889	28.25833	From Kolobe camp in Lapalala; drive out through the back road. Go down a

									steep concrete road; take the track to the West down into the Bloklandspruit river- cross small stream. At the T-junction turn left cross Main River. Drive 400m and park. The cave is half way up the cliff across the river.",
Schroda VI 46 MS	ZA	Limpopo	Polokwane	Onverwacht	0530	Messina (Musina)	-22.21111	29.41389	null
Athens II 31 MS	ZA	Limpopo	Polokwane	Beitbridge	0900	Soutpansberg	null	null	null
Little Leigh I 730 MS	ZA	Limpopo	Polokwane	Tshikuwi	0900	Soutpansberg	null	null	null
Pont Drift I 12 MS	ZA	Limpopo	Polokwane	Onverwacht	0900	Soutpansberg	null	null	null
Hetty I 93 MT	ZA	Limpopo	Polokwane	Tshiungani	0530	Venda (Thohoyandou)	null	null	null
Den Staat I 27 MS	ZA	Limpopo	Polokwane	Onverwacht	0530	Soutpansberg	null	null	Start on Den Staat road. Turn on to Den Staat farm at the private road

									sign. Take the first road to the right; follow it to the end (there is a large baobab and a khami ruin). Walk south from here over rocky area. The site is on the edge of the rocky outcrop, where it drops down to the plane.",
Too Late XXVIII 359 LR	ZA	Limpopo	Polokwane	Onverwacht	0709	Bochum	null	null	null
Too Late XXXIX 359 LR	ZA	Limpopo	Polokwane	Onverwacht	0709	Bochum	null	null	null
Doppie III 95 MT	ZA	Limpopo	Polokwane	Onverwacht	0950	Venda (Thohoyandou)	null	null	null
Alicedale II 138 MT	ZA	Limpopo	Polokwane	Onverwacht	0900	Messina (Musina)	null	null	null
Amonda III 161 MT	ZA	Limpopo	Polokwane	Onverwacht	0900	Messina (Musina)	null	null	null
David VI 160 MT	ZA	Limpopo	Polokwane	Onverwacht	0900	Messina (Musina)	null	null	null
Dawn I 71 MT	ZA	Limpopo	Polokwane	Onverwacht	0900	Messina (Musina)	null	null	null

Doreen I 108 MT	ZA	Limpopo	Polokwane	Onverwacht	0900	Messina (Musina)	null	null	null
Schroda XVI 46 MS	ZA	Limpopo	Polokwane	Onverwacht	0900	Messina (Musina)	null	null	null
Millstream VIII 358 LR	ZA	Limpopo	Polokwane	Onverwacht	0699	Seshego	null	null	null
Edmondsburg II 32 MS	ZA	Limpopo	Polokwane	Onverwacht	0900	Soutpansberg	null	null	null
Hackthorne V 30 MS	ZA	Limpopo	Polokwane	Onverwacht	0900	Soutpansberg	null	null	null
Mont Blanc LXII	ZA	Limpopo	Polokwane	Driekoppies and Gamonyebodi	0709	Bochum	-23.29289	28.78194	null
De Villiersdale I	ZA	Limpopo	Polokwane	Driekoppies and Gamonyebodi	0709	Bochum	-23.26531	28.79119	200m North-East of site two.
Diepkloof I 41 LS	ZA	Limpopo	Polokwane	Muungadi and Tshikwarani	0709	Soutpansberg	-23.01417	29.40556	null
Doornleegte I 594 LR	ZA	Limpopo	Polokwane	Onverwacht	0709	Waterberg	-23.90833	28.34167	Take the track from the hikers tent camp downstream Lapalala River. For about 1.5km. Then scale down from the high point on the cliff and the site is just under an overhang

									face. It is a small and well sheltered site.",
Rynie I 158 MT	ZA	Limpopo	Polokwane	Musekwa	0900	Messina (Musina)	-22.63389	30.2425	null
Haakdoorn Draai I 711 LR	ZA	Limpopo	Polokwane	Ga Mathekga	0746	Mokerong 2	-23.6375	28.57083	Take road from Potgietersrus to Marken. From Marken turn right (East) at the T-junction. The road is tar for few hundred meters, and then becomes dirt. Drive till road turns to the right (South), you will see a sign to Masebe Game Reserve. From Masebe reserve gate take the gate opposite to what is called \"A\" reserve. The gate you take is smaller than the main one. Follow the

									road onto the reserve for about 200m. The road goes left and curves up to the rock art site.",
Machete I 29 MS	ZA	Limpopo	Polokwane	Onverwacht	0900	Soutpansberg	-22.24994	29.2925	null
Little Muck II 26 MS	ZA	Limpopo	Polokwane	Vergenoeg	0728	Soutpansberg	-22.24556	29.25972	null
Samaria III 28 MS	ZA	Limpopo	Polokwane	Onverwacht	0900	Soutpansberg	-22.20975	29.32014	null
Zoutpan I 459 MS	ZA	Limpopo	Polokwane	Kromhoek	0709	Soutpansberg	-22.975	29.33194	null
De Villiersdale III	ZA	Limpopo	Polokwane	Driekoppies and Gamonyebodi	0709	Bochum	-23.25278	28.80417	null
Mont Blanc IX	ZA	Limpopo	Polokwane	Driekoppies and Gamonyebodi	0709	Bochum	-23.26806	28.825	null
Mont Blanc X	ZA	Limpopo	Polokwane	Driekoppies and Gamonyebodi	0709	Bochum	-23.26333	28.82083	null
Mont Blanc XI	ZA	Limpopo	Polokwane	Driekoppies and Gamonyebodi	0709	Bochum	-23.25917	28.82	null
Bulsfontein III	ZA	Limpopo	Polokwane	Onverwacht	0500	Ellisras	-23.99667	27.53917	Take the old Ellisras road from Vaalwater, pass Bulgerivier. Turn left at the

									sign to Steenbokpan. 11.5 km down the road there is a sleeper on the left saying \"Kurumakatiti\". Directions to site can be obtained from the farm owner.",
Baviaanskrans III 659 LR	ZA	Limpopo	Polokwane	Mongatane and Gamusi	0617	Mokerong 2	-23.49722	28.52583	Approximately 100m from Baviaanskrans II- Nearly directly opposite the Ga Hlala settlement.
Haakdoorn Draai IV 711 LR	ZA	Limpopo	Polokwane	Ga Mathekga	0617	Mokerong 2	-23.63833	28.57056	When facing Haakdoorn Draai I follow the rock face to the right (West). The site will be located at approximately 60m.
Frishgewaagd III 590 LR	ZA	Limpopo	Polokwane	Ga Mosidi	0650	Potgietersrus	-23.91644	28.34978	Just north of Frishgewaagd II. Higher up on the cliffs.

Bonne Esperance VII 356 LR	ZA	Limpopo	Polokwane	GaMonyebod i	0699	Seshego	-23.29389	28.82917	null
Bonne Esperance VIII 356 LR	ZA	Limpopo	Polokwane	GaMonyebod i	0699	Seshego	-23.29606	28.831	null
Bonne Esperance XIV 356 LR	ZA	Limpopo	Polokwane	GaMonyebod i	0699	Seshego	-23.29603	28.83206	null
Bonne Esperance XIX 356 LR	ZA	Limpopo	Polokwane	GaMonyebod i	0699	Seshego	-23.29056	28.84111	null
Bonne Esperance XXX 356 LR	ZA	Limpopo	Polokwane	GaMonyebod i	0699	Seshego	-23.29306	28.80833	null
Tromp I 252 LS	ZA	Limpopo	Polokwane	Louis Trichardt	0920	Soutpansberg	-23.00972	29.78417	null
Dwaalhoek IV 185 KQ	ZA	Limpopo	Polokwane	Vaalwater	0530	Waterberg	-24.06944	27.81944	About 15m from Dwaalhoek III. For directions see Dwaalhoek I.
Graaflust III 637 LR	ZA	Limpopo	Polokwane	Onverwacht	0530	Waterberg	-23.94861	28.25278	Just above Graaflust II in the next level of the rock.
Graaflust IV 637 LR	ZA	Limpopo	Polokwane	Onverwacht	0530	Waterberg	-23.94722	28.25417	Take road to Graaflust II instead of crossing the river, walk along the bank to theright. Cross the river where

									there are many black rocks. In front there is an obvious shelter.
Kliphoek IV 636 LR	ZA	Limpopo	Polokwane	GaMosidi	0530	Waterberg	-23.91944	28.25889	Approximately 80m along the same cliff line from Kliphoek III.
Klipplaat IIIC 34 KR	ZA	Limpopo	Polokwane	Vaalwater	0530	Waterberg	-24.04861	28.30611	null
New Belgium III 608 LR	ZA	Limpopo	Polokwane	Vaalwater	0530	Waterberg	-23.83611	28.25278	Drive to Lapalala West bush camps. Follow the signs to Mukwa camp. Cross the river at the rocks and walk up the river. Just before the bend in the river, there is a shelter on your left.
Alem I 544 LR	ZA	Limpopo	Polokwane	Vaalwater	0530	Waterberg	-23.78583	28.27083	null
Schroda II 46 MS	ZA	Limpopo	Polokwane	Vaalwater	0530	Messina (Musina)	-22.18111	29.42833	null
Schroda III 46 MS	ZA	Limpopo	Polokwane	Vaalwater	0530	Messina (Musina)	-22.1825	29.42917	null
Little Muck I 26 MS	ZA	Limpopo	Polokwane	Vergenoeg	0900	Soutpansberg	-22.24556	29.25972	null

Folorodwe IV 79 MT	ZA	Limpopo	Polokwane	Onverwacht	0950	Venda (Thohoyandou)	-22.51778	30.49861	null
Nieuwe Jerusalem I 327 LR	ZA	Limpopo	Polokwane	Drikoppies and Gamonyebodi	0709	Bochum	-23.25472	28.86583	null
Palmietfontein I 32 KQ	ZA	Limpopo	Polokwane	Vaalwater	0530	Waterberg	-24.04861	28.33278	East side of main road, opposite cattle lands farm sign. Follow track to Melkrivier valley. Cross river, follow yellow paint splodges, 30m up on the West side of the riverbank. Shelter hidden by trees.
Preezburg I 400 LR	ZA	Limpopo	Polokwane	Monte cristo	0617	Mokerong 2	-23.43472	28.70556	Take track running through Preezbury from either side. At the river drift look at the north bank for a cattle trough (cement). Turn along a small track running north,

									on the left there is a wind pump; continue until up against the hillside. The site is on the boulder on the corner of the hill as it turns on the bulge of the hill. The site is where two boulders lean into one another.
Waterval I 601 LQ	ZA	Limpopo	Polokwane	Onverwacht	0555	Ellisras	-23.85694	27.66111	Take the old road between Vaalwater and Ellisras, approximately 25kms from Ellisras. Just before the bend on the road a large cave can be seen on the left hand side of the road. Just before one is directly opposite the cave, there is a

									small overgrown road (on the left), leading to abarbed wire and stick gate. Follow the road, which later becomes a path towards the river, cross the river and head up to the cave.",
Buffelskom I 764 MS	ZA	Limpopo	Polokwane	Onverwacht	0900	Soutpansberg	-23.00833	29.19417	null
Bulsfontein I	ZA	Limpopo	Polokwane	Onverwacht	0555	Ellisras	-23.99514	27.53917	Take the old Ellisras road from Vaalwater, pass Bulgerivier. Turn left at the sign to Steenbokpan. 11.5 km down the road there is a sleeper on the left saying \"Kurumakatiti\". Directions to site can be obtained from the farm owner.",

Dambale I	ZA	Limpopo	Polokwane	Domboni and Mataulu	0950	Venda (Thohoyandou)	-22.49889	30.73111	null
Den Staat II 27 MS	ZA	Limpopo	Polokwane	Domboni and Mataulu	0900	Soutpansberg	-22.23194	29.24389	Start on Den Staat road. Turn on to Den Staat farm at the private road sign. Take the first road to the right; follow it to the end (there is a large baobab and a khami ruin). Walk south from here over rocky area. The site is on the edge of the rocky outcrop, where it drops down to the plane.",
Graaflust I 637 LR	ZA	Limpopo	Polokwane	Domboni and Mataulu	0530	Waterberg	-23.92639	28.2175	From Vaalwater take the road to Melrivier (right) for about 32 km, passing Nyathi Game Lodge on your right.

									You will then come across a left turning, with a sign to Betesda Farm and Dorset. The road is tarred for a short while, and then becomes dirt. Follow this road for about 15km(don't turn left again). You will then see the Rivierdans Farm sign on the right hand side of the road. From the farm take the road heading right just before the start of the farmhouse, and then you need to ask the owners for the directions to the site.",

Groothoek I 220 KR	ZA	Limpopo	Polokwane	Potgietersrus	0600	Potgietersrus	-24.12917	28.65389	null
Witbank I 647 LQ	ZA	Limpopo	Polokwane	Onverwacht	0555	Ellisras	-23.98889	27.71944	The site is located approximately 250m South of the Dam wall on the right. The small cave is quite obvious from the water level and it is fenced off at the moment. Since the Dam wall is out of bounds to the unauthorized, it is better to get to the cave by boat.
De La Roche VIII	ZA	Limpopo	Polokwane	Gamonyebodi	0709	Bochum	-23.28778	28.79667	null
De Villiersdale II	ZA	Limpopo	Polokwane	Driekoppies and Gamonyebodi	0709	Bochum	-23.26731	28.78989	null
Samaria IV 28 MS	ZA	Limpopo	Polokwane	Domboni and Mataulu	0900	Soutpansberg	-22.22058	29.33653	This site is on the same hill as Samaria I. It is on the western side in a small

									reserved sheltered area.
Haakdoorn Draai V 711 LR	ZA	Limpopo	Polokwane	Ga Mathekga	0617	Mokerong 2	-23.62833	28.56722	null
Groothoek II 220 KR	ZA	Limpopo	Polokwane	Potgietersrus	0650	Potgietersrus	-24.12361	28.64333	null

Appendix D: Ethical clearance approval letter

School of Geography, Archaeology and Environmental Studies

Private Bag 3, Wits 2050, South Africa Enquiries: GEOGRAPHY: TEL: +27 11 717-6503 • FAX: +27 086 651 6366
ARCHAEOLOGY: TEL: +27 11 717-6045 • FAX: +27 086 651 6366
http://www.wits.ac.za/geography/

15 January 2019

Internal Ethics Clearance Certificate

Name and number: **Bellina Chokoe 1727895**

Degree: **MSc coursework**

Title of Project: **Rock art in the geographical context: the design of a spatial database for the South African Rock Art Digital Archive (SARADA)**

Protocol number: **GEOG-5-2019-0001**

Valid until: **December 2020**

Dr Alex Wafer
Ethics Committee Chair, Geography Division
School of Geography, Archaeology and Environmental Studies

Declaration by the Investigator(s) (copy to be returned to the chair of the committee)

I / we acknowledge that this clearance certificate has been granted on the basis of the project described above and submitted to the research ethics committee (internal). I / we acknowledge that the ethics certificate is granted in compliance with the University of the Witwatersrand Research Ethics (non-medical) protocol, and that as such I am bound by that protocol. Any changes to the above-mentioned research project must be communicated to the ethics committee for approval.

Appendix E: Positional error measure

ID	Name	Y_START	X_START	Y_END	X_END	Shape_Leng	Distance(m)
0	Witbank I 647 LQ	-23,9889	27,71806	-23,9889	27,71944	0,001377715	140
1	Fancy I	-23,8793	27,63761	-23,8793	27,63769	8,24838E-05	8
2	Waterval I 601 LQ	-23,857	27,66112	-23,8569	27,66111	0,000102643	11
3	RSA CRD1	-23,6708	27,93068	-23,6708	27,93056	0,000139139	14
4	Alem I 544 LR	-23,7863	28,27096	-23,7858	28,27083	0,000514723	57
5	New Belgium III 608 LR	-23,8361	28,25278	-23,8361	28,25278	3,21407E-05	4
6	New Belgium II 608 LR	-23,8417	28,26615	-23,8417	28,26611	4,6686E-05	5
7	New Belgium I 608 LR	-23,845	28,26808	-23,845	28,26806	5,18998E-05	6
8	Graaflust I 637 LR	-23,9263	28,2176	-23,9264	28,2175	0,000114359	12
9	Kliphoek IV 636 LR	-23,9191	28,25892	-23,9194	28,25889	0,000312158	35
10	Kliphoek I 636 LR	-23,9056	28,25553	-23,9056	28,25556	3,27806E-05	3
11	Graaflust IV 637 LR	-23,9473	28,25417	-23,9472	28,25417	8,30393E-05	9
12	Graaflust II 637 LR	-23,9487	28,25416	-23,9486	28,25417	8,79081E-05	10
13	Graaflust III 637 LR	-23,9488	28,25273	-23,9486	28,25278	0,000168604	19
14	Sliedrecht II 638 LR	-23,9533	28,25887	-23,9533	28,25889	2,45285E-05	3
15	Graaflust VI 637 LR	-23,9499	28,26947	-23,95	28,26944	0,000100812	11
16	Klipplaat II 34 KR	-24,0487	28,3125	-24,0486	28,3125	6,08223E-05	7
17	Klipplaat III 34 KR	-24,0486	28,31398	-24,0486	28,3125	0,00147576	150
18	Palmietfontein I 32 KQ	-24,0486	28,33277	-24,0486	28,33278	5,93997E-06	1
19	Olievenfontein II 111 KR	-24,1585	28,40182	-24,1583	28,40194	0,00024295	26
20	Olievenfontein I 111 KR	-24,1567	28,40232	-24,1569	28,40278	0,000517854	54
21	Groothoek I 220 KR	-24,1298	28,65459	-24,1292	28,65389	0,000942628	100
22	Mont Blanc XIII	-23,2773	28,86104	-23,2758	28,85833	0,003082045	321

23	Mont Blanc X	-23,2608	28,82547	-23,2633	28,82083	0,005262487	549
24	Nieuwe Jerusalem I 327 LR	-23,2577	28,86924	-23,2547	28,86583	0,004521641	480
25	De La Roche VIII	-23,2879	28,79855	-23,2878	28,79667	0,001884564	193
26	Balerno I 18 MS	-22,2875	29,20195	-22,2875	29,20194	2,50612E-05	3
27	Balerno II 18 MS	-22,2875	29,20278	-22,2875	29,20278	7,00209E-06	1
28	Armenia VII 20 MS	-22,2792	29,21501	-22,2792	29,215	1,77353E-05	2
29	Armenia XI 20 MS	-22,2789	29,21528	-22,2789	29,21528	1,19218E-05	1
30	Armenia V 20 MS	-22,272	29,22945	-22,2719	29,22944	1,57711E-05	2
31	Armenia IV 20 MS	-22,2717	29,22944	-22,2717	29,22944	4,34264E-06	0
32	Armenia VI 20 MS	-22,27	29,22972	-22,27	29,22972	3,28184E-06	0
33	Balerno IV 18 MS	-22,265	29,18806	-22,265	29,18806	4,20119E-07	0
34	Modena I 13 MS	-22,2544	29,17375	-22,2542	29,17378	0,000185201	20
35	Modena II 13 MS	-22,2541	29,15784	-22,2539	29,15778	0,000228747	25
36	Rhodes Drift I 22 MS	-22,2278	29,18432	-22,2274	29,18419	0,000501784	55
37	Little Muck II 26 MS	-22,2456	29,25989	-22,2456	29,25972	0,000168207	17
38	Den Staat II 27 MS	-22,232	29,24391	-22,2319	29,24389	4,84215E-05	5
39	Machete I 29 MS	-22,25	29,29266	-22,2499	29,2925	0,000165677	17
40	Athens I 31 MS	-22,2592	29,33625	-22,2592	29,33611	0,000143057	15
41	Samaria V 28 MS	-22,206	29,33742	-22,2039	29,33653	0,002280256	250
42	Samaria IV 28 MS	-22,2208	29,33709	-22,2206	29,33653	0,000588103	61
43	Greefswald X 37 MS	-22,2135	29,35518	-22,2136	29,35472	0,000477208	49
44	Greefswald II 37 MS	-22,2167	29,35436	-22,2167	29,35389	0,000471142	49
45	Greefswald VIII 37 MS	-22,2247	29,36862	-22,2253	29,36833	0,000636647	70
46	Greefswald I 37 MS	-22,2314	29,37446	-22,2314	29,37444	2,79963E-05	3

47	Greefswald VII 37 MS	-22,2301	29,37039	-22,2306	29,37028	0,00044109	49
48	Greefswald XII 37 MS	-22,2335	29,37425	-22,2334	29,37394	0,000307976	32
49	Schroda VI 46 MS	-22,2109	29,41406	-22,2111	29,41389	0,000284703	31
50	Dambale I	-22,4879	30,71499	-22,4989	30,73111	0,019500738	2059