ARID ZONE JOURNAL OF ENGINEERING, TECHNOLOGY & ENVIRONMENT
AZOJETE September 2022. Vol. 18(3):541-554
Published by the Faculty of Engineering, University of Maiduguri, Maiduguri, Nigeria.
Print ISSN: 1596-2644, Electronic ISSN: 2545-5818
www.azojete.com.ng

ORIGINAL RESEARCH ARTICLE

EXTRACTION AND GEO-REFERENCING SUBURBAN LOCATIONS OF TERRORISTIC ATTACKS FROM TEXTUAL DATA OF MAIDUGURI INSURGENCY

S. S. Garba
Department of Geomatics, University of Maiduguri, Maiduguri, Nigeria
**Corresponding author's email address: ssgarba@unimaid.edu.ng*

ARTICLE INFORMATION

Submitted 21 May, 2022
Revised 22 August, 2022
Accepted 31 August, 2022

Keywords:
Coordinates
Fuzzy Lookup
Location Extraction
Georeferencing
Reference Location Data
Suburban and Terrorism Attack

ABSTRACT

There are no precise suburban coordinates of location of terrorism attacks in parts of Maiduguri. Effort to apply geospatial analysis will require such data. The objective of this study was to enforce georefencing to textual global terrorism databases, particularly for Maiduguri, such that it will provide basis for geospatial analysis and planning counter terrorism. Global terrorism data were sourced from Internet and reduced to Maiduguri and its surroundings. The textual aspect of the databases that reported attacks were used to extract the phrases containing prepositions that indicated location of the attacks at suburban level (such as wards). The phrases were cleaned in order to bring out the names of the locations of attacks. A reference location data (RLD) was developed from OpenStreetMap (OSM) and the Nigerian Administrative boundary (SURCON, 2022) and used in determining the coordinates of the locations using Microsoft Excel Fuzzy look up. Half of the extracted names could not be determined from the OSM based RLD. The names that could not be found were looked up and determined from the Google Earth. The preposition 'in' and 'at' were the most useful, returning most of the extracted location. Other indicators of locations such as: ward, area and neighbourhood merely repeated what the two prepositions have determined. The process extracted 188, 273 and 169 geospatially records for Global Terrorism Database (GTD), Gedevents and Armed Conflict Location and Event Data Project (ACLED) respectively. This work demonstrated that global data on terroristic attack can be used to determine suburban location of attacks. Conversely, tracking terroristic attacks should include suburban locations. The key to determining the coordinates of a suburban location is a robust reference dataset, which can be updated. This work provided the basis for geospatial analysis and hence counterterrorism efforts.

1.0 Introduction

Maiduguri is the location where the Boko Haram Insurgency began and persisted in the North Eastern Nigeria from 2009 up to the present time. There are continuous collection and archiving of reports on the conflict especially by the Nigeria and foreign media such as the Agence France-Presse (AFP). There are also dedicated efforts to track global terrorism with their spatial characteristic in order to give insight into the process of terrorism (Bahgat and Medina, 2013).

Fundamentally, reports on terroristic attacks are not concerned with the geospatial precision such as the coordinates but rather the names of location. The media (Newspapers) generally reported terroristic attacks as happening 'in Maiduguri' (which may be sufficient for global audience), but such use of names lacks geospatial precision. However, the media occasionally referred to particular location of suburban areas of Maiduguri such as: area, ward, popular name of a place or neighbourhood. The mentioning of the parts of the town provides the opportunity for determining the spatial characteristics of the location. These suburban names are often buried in long paragraphs of the reports or datasets. Hence this research is focussed on the problem of extracting the names of the locations from the paragraphs and georeferencing them. This aimed at understanding and mapping location of terroristic attacks that were originally recorded without precise geospatial properties. In this study the various questions of why a terrorism happened at a particular location or how or who were involved and the effects are out of the scope of this work.

Effort to relate textual material to geographical coordinates had received attention across such disciplines like dialectology, computational sociolinguistics and Geographical Information Retrieval (GIR), geospatial semantic Web and Geonames ontology (Melo and Martins, 2017). Advanced efforts in detecting toponyms such in Radke (2019), that assumes reference data already existed, hence propositions with a test are used to automatically determine a the toponyms from the existing reference. However, the extraction of specific suburban name from public names and data is complex matter (Melo and Martins, 2017).

The research problem can be defined formally in the word of Skoumas *et al.* (2016): "Given a set of objects PV with a-priori known coordinates in space, a set of objects PU whose exact positions are unknown and a set of predefined spatial relationships R between PU and PV objects, find probabilistic estimates for the positions of PU objects in space."

In this study, *PV* is a set of locations in Maiduguri: names of wards, districts, streets, roads and place whose coordinates are known or can be located. *PV* is the reference location dataset (RLD). *PU* is a set of all extracted location of attacks from media material that are not georeferenced. The aim of this study was to match extracted set of location, to existing location that has precise geospatial properties. It is also to show the complexity of such endeavour as it relates to Maiduguri Township and develop guide relating to the rules of the relationship between them.

2.0 Methodology

There are four components of the method as shown in Figure 1. The first part consists of collecting the data and extracting the textual description of location of terrorist attract. The second part involves the development of a reference location database which will be used in determining the coordinates of the extracted locations. The third was the determination of the coordinates. The fourth was the culmination of the process, which is a Geospatial database (including maps). A provision is made so that when the reference database fails to provide coordinates because the given location is not in reference location database; the locations are checked in Google Earth Pro. The product from the Google earth becomes input to the Reference location database.

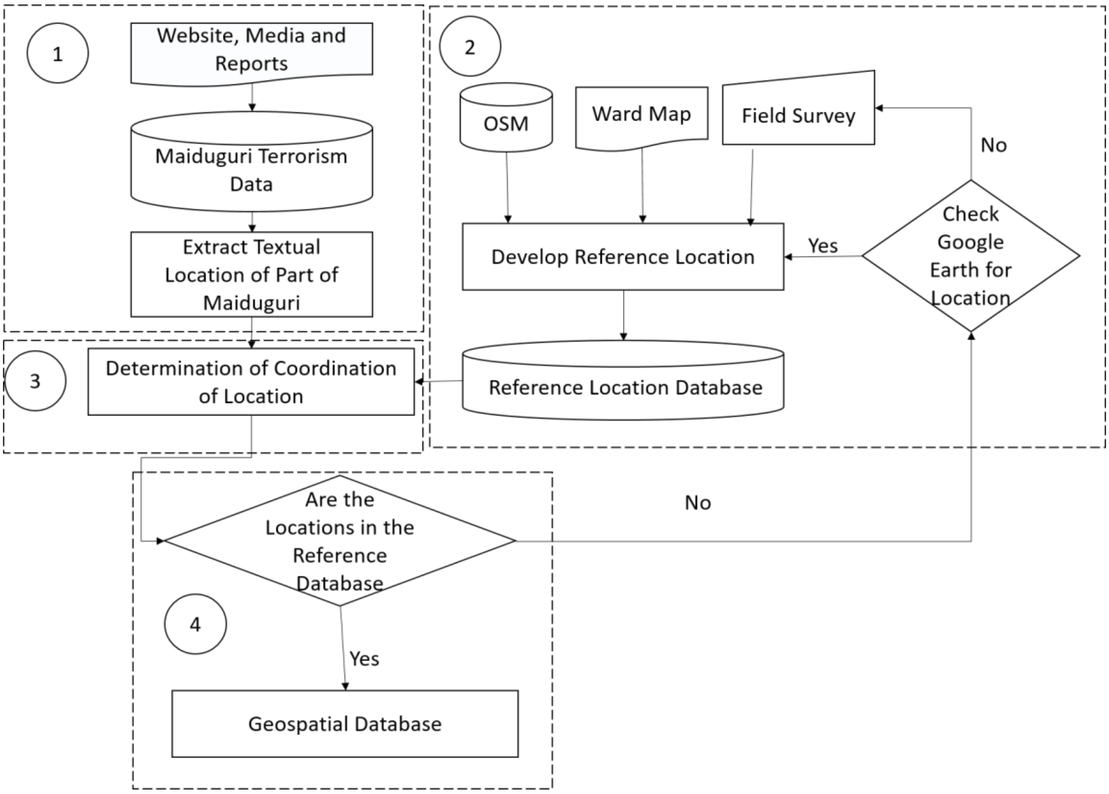

Figure 1: Process of Georeferencing extracted location and database development. (1) the sources of data and the extraction process, (2) the reference data set with geospatial characteristics developed for Maiduguri area (3) the georeferencing and (4) the product of the process.

1.1 The Extraction of Textual Location of Terroristic Attack

Below was the process used in extracting names of locations where terroristic attack happened in Maiduguri from database of global terrorism tracking. This subsection explains the tracked data and the process of extraction of names.

1.1.1 Globally Tracked Terrorist Data

Data on terroristic attacks exist on the Internet, they are generally given from global or regional perspective, it is from datasets that the Maiduguri information was extracted. This study was particularly concerned with the data relating to Maiduguri during the period of insurgency from 2009 to 2020. Three global datasets on terrorism were used: The Armed Conflict Location and Event Data Project (ACLED); Global Terrorism Database, (GTD, University of Maryland, 2022) and the Uppsala Conflict Data Program (UCDP). The ACLED began as a PhD research work with focus on conflict in Africa. Its first version was released in 2009 and the current version is the 8th, released in 2018. The GTD is maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START) at the University of Maryland. It began in 2001 to work on archiving data of Pinkerton Global Intelligence Services (PGIS). The work was expanded in the type of data and definition and integration of data from the Institute for the Study of Violent Groups (ISVG) (University of Maryland, 2022). The UCDP is a project under the Department of Peace and Conflict Research, Uppsala University. It is continuously updated online database on armed conflicts and organised violence (Uppsala University, 2022). The volume of records, meaning the number of incident reported and the associated aspects of the record Heading or Fields are given in Table 1. Since the downloaded data comes in Microsoft Excel format, the records are the rows of the Excel Sheet and the categories are the columns.

Table 1: Volume of original downloaded data

Database	No of Records	No of Heading (Fields)	Period
ACLED	20614	32	07/01/2009 to 01/01/2022
GTB	5070	135	13/2/1976 to 29/12/2019
UCDP	4604	47	25/9/1990 to 31/12/2019

All the databases obtained their data by extracting from Major international news agencies and Nigerian media outfits. Table 2 provides the summary of the sources according to: International News Sources, Nigerian Newspapers and Social Media. Since the downloaded data cover the whole of Nigeria, it became necessary to reduce them to Maiduguri and environs and between Latitude (11.780627°N and 11.889901°N) and Longitude (12.975890°E and 13.261794°E) and from 2009 – 2019.

Table 2: Sources of data of the databases

Database	International News Sources	Nigerian Newspapers	Social Media	Remark
ACLED	28	32	1	The most used media are the Agence France-Presse (AFP) and Nigerian Daily Trust and Guardian
GTB	50	21	Nil	The most used media are Agence France-Presse (AFP) and Nigerian Daily Trust, Vanguard and Guardian
UCDP	43	55	4	The most used media are the All African website, Agence France-Presse (AFP) provided the international sources, the Nigeria Watch. It also sourced from Human Right Watch (HRW) and Twitter.

1.1.2 The Extraction of Location of attack

The extraction process involved, searching and removing text or phrase that indicated location of attack from the column containing the textual reports of the attacks. Extraction processes assumes that certain prepositions or word will indicate where an attack happened. It also assumes that there are not enough reference names to use to search media materials directly, and the familiarity of the researcher confirms the names. The column containing the reports of the incidents was used for the search and extraction. The extraction was conducted in four steps.

The first step was to extract phrase containing the following prepositions: ‘in’, ‘at’, ‘along’, ‘on’ and ‘near’. Each was extracted separately. The process was conducted by formulating the

Microsoft Excel to pick a phrase associated with a preposition of interest. An example of the process is given in Table 3. The Event column reports an attack on the 21st September 2010 from the GTD data. From the report which in cell C12 of the Excel sheet, the formula will search for the preposition "in" and give the resulting phrase containing forty letters containing space.

Table 3: Example of Microsoft Excel Formula to extract phrase with the word "in"

Event	Formula
09/21/2010: On Tuesday, in the Gwaidomari neighborhood i, Borno, Nigeria, militants riding motorcycles fired upon and killed two people with Kalashnikov style rifles. One local chief and one trader were killed. A man claiming to be a chieftain of Nigerian Taliban, also known as Boko Haram, claimed responsibility on behalf of the group.	=MID(C12,(SEARCH("in",C12,1)+3),40) Result the Gwaidomari neighborhood in, Borno, Ni

The second step was the determination of the actual location name from the results of first step based on the researcher's knowledge of the place. The third step, was to remove duplication of results from the prepositions. Finally, extraction was conducted by using area related words such as district, area, wards and neighbourhood and removing duplication with result of the third steps. These process were used for the three datasets.

1.2 Development of Reference Location Dataset (RLD)

RLD refers to data set containing names of places within Maiduguri and environs with their coordinates which will be used to determine the coordinates of the extracted locations. The data was built using the ward map of Nigeria (OSGOF, 2022), OpenStreetMap (OSM, Haklay and Weber, 2008) and Google Earth Pro. The ward map provided the traditional names of areas, most likely association with areas at sub of a Local Government Areas or a geographical organisation of a chiefdom. The Maiduguri Urban encompasses three local government areas (Metropolitan, Jere and Konduga), it has 27 wards aligned to Independent National Election Commission (INEC) distribution, which evolved from two districts (Waziri, 2012). Over the years some wards have split into two or three such as Gwange, Bolori and Shehuri (with suffix I, II, III. North and South). The wards are areas however they are represented by the coordinates of their geometric centres. The wards that have split over the years presented a problem whenever they are mentioned without their suffix. For example, if a report mention only Gwange, it will be impossible to know which of the three Gwange. Hence a super ward was created that combines all wards with similar origin and the coordinates of the centroid was used.

OSM provided locations more precise location of places other than the wards. It has eight layers of georeferenced information: Places, Points, Railways, Roads, Waterways, Buildings, Land use and Natural. The RLD was further expanded whenever an extracted name of a

Corresponding author's e-mail address: ssgarba@unimaid.edu.ng

location was not found in the RLD. Such name was then checked up (Start up the software and search) in the Google Earth Pro (and its coordinates determined and then added to the RLD.

1.3 Georeferencing: Determination of Coordinates of Extracted Names

The extracted names from the three datasets were each looked up in the RLD in order to determine their coordinates. This was what amounted to georeferencing of the extracted textual name of location. The Microsoft Excel Fuzzy Lookup (Microsoft, 2022) was used to do search and match up. The software is powerful in automating the search and match and tolerance of misspelling.

1.4 Predefined Relationships

The formal definition of the research problem, requires a define spatial relationship. This study does not go into Geospatial Information Systems (GIS) topology, but focussed on the relationship between extracted textual and RLD. Some the rules are:

1. That extracted names that matched ones in the RLD will adopt the coordinates of the ones in the RLD.
2. That closely associated names arising due to small different in spelling or misspelling of the names or the short form the names should be accommodated (hence the use of Microsoft Excel Fuzzy Lookup).
3. That RLD should be built to accommodate fuzziness.

2. Results and Discussion

2.1 The Reduction of Maiduguri Data

The downloaded datasets were reduced by two criteria: the name Maiduguri and by given coordinates and summarised in Table 4. The summary shows that selection by latitude and longitude provides greater record of event of attacks than by simply using the name Maiduguri. There are more records from ACLED than the other dataset.

Table 4: Number of records selected by either Name or Coordinate

	Selection by Name	Selection by Latitude and longitude
Gedevents	392	423
Gtd_Nigeria	493	557
ACLED	680	737

2.2 The Extraction of Data

It is viable to use location associated word, namely: in, at, along, on, near, neighbourhood, ward, district and area in order to search and extract phrase containing the locations of attacks from GTD and ACLED dataset. The results such process were summarised in Table 5. The preposition ‘in’ provided the highest result (557 of 596 of extraction from GTD), however not every result with the ‘in’ gives a location of part of Maiduguri. Many of the ‘in’ are associated with ‘in Maiduguri’. Hence the cleaning, which reduced half of the results in the GTD dataset and reduced the records of the ACLED to 20%. The ‘at’ provide the second highest. The cleaning was necessary because many of at are associated with time such as ‘at 3pm’. On cleaning the result of GTD was reduced to 25% and 35% of ACLED. The repeat column of Table 5 indicates whether a word was repeating the extraction of another, thus in the GTD data, the ‘at’ is not repeating ‘in’ and in the ACLED only 5 of 61 are repeating the ‘in’. The ‘along’, ‘on’ and ‘near’ all reported very few results. The neighbour, district, area and ward reported high results but they are mostly reporting what the ‘in’ and ‘at’ had reported. Eventually 389 records were extracted from the GTD and 255 from the ACLED. The process of cleaning the data relied on the local knowledge of names.

Table 5 Prepositional Performance

Definer	Dataset			
	GTD	Repeated	ACLED	Repeated
in	557		444	
in (Cleaned)	251		94	
at	310		176	
at (cleaned)	75		61	5
Along	8	3	35	3
on	276		208	
on (cleaned)	11	3	31	7
near	53		64	
near	36	34	30	30
neighbourhood	41	41	10	3
ward	36	36	23	8
district	17		2	2
area	121		111	
Area (Cleaned)	121	90	83	56
Total Cleaned	596	207	369	114
Actual (less Repeated)	389		255	

The third dataset the GEDEVENTS data does not have a summary of events that could be used in extraction of data as the other two. However, it had a column called “where description” in which the 423 records which reported 258 (61%) locations of part of Maiduguri.

The process of extraction shows that most of the records from GTD provided single name of location but ACLED and GTDEVENT have multiple names per record. The multiple record implied the case where on the date there were multiple attacks in different parts of Maiduguri. And sometimes it was simply the description of one event in a larger area, such as a point in a ward. The extracted data are sometime repeated because an attack was repeated at a particular location. Tables 6, 7 and 8 are excerpt from the extracted results of GTD, ACLED and Gedevents respectively. it could be seen that names may not always a single, such as Aladuwari (Table 7) but complex such as 333 Artillery Nigerian Army base/market (Table 6).

Table 6 Excerpt from GTD,

eventid	Date	summary	Extracted
201806200021	06/20/2018	Two suicide bombers detonated at the 333 Artillery Nigerian Army base/market gate i, Borno, Nigeria. The two assailants were killed and at least 15 people were injured in the blasts. No group claimed responsibility	333 Artillery Nigerian Army base/market
201506020031	06/02/2015	A suicide bomber detonated at a cattle market i city, Borno state, Nigeria. In addition to the bomber, at least 31 people were killed and 25 people were injured in the blast. No group claimed responsibility for the incident	Cattle market

Table 7 Excerpt from the ACLED Extracted

event_id_no_cnty	event_date	notes	Extracted
1306	27-Dec-15	Nigerian forces shoot and kill seven potential suicide bombers (some reports claim 14) attempting to gain access to Maiduguri (at Aladuwari area) on 27/12 This follows the detonation by three of bombers at a nearby mosque, killing 25 and injuring 85. The bombers were reported to be teen-aged girls.	Aladuwari
4429	04-Dec-21	On 4 December 2021, ISWAP militants claimed to have fired 6 mortar shells at the 1000 residential housing estate, Gomari and 777 housing estate in Maiduguri (Maiduguri Metro, Borno). At least one house was damaged and one baby was wounded, the rockets were fired from the Duwari area. Other sources report that soldiers engaged the militants in an armed clash on the outskirts of Maiduguri.	1000 residential housing

Table 8 Excerpt of Gedevents Extraction

id	where_description	Extracted	date_start	date_end
258504	the informal market at the base of 333 Artillery Nigerian Army in Maiduguri	333 Artillery Nigerian Army	04/08/2012	04/08/2012
269489	an area called Caterpillar near Ajilari cross, Airport Area in Maiduguri	Caterpillar near Ajilari cross, Airport Area in	04/10/2012	04/10/2012
268150	on the outskirts of Maiduguri, attacking an IDP camp in Judumri area, not far away from the Federal High Court complex, Jiddari area	Federal High Court complex, Jiddari area	04/26/2012	04/26/2012
255846	on the outskirts of Maiduguri, attacking an IDP camp in Judumri area, not far away from the Federal High Court complex, Jiddari area	Judumri	06/05/2012	06/05/2012
284313	in Muna Dalti on the outskirts of Maiduguri	Muna Dalti	06/17/2012	06/17/2012

2.3 The Reference Location Data (RLD)

Reference data were first developed from the thirty-three wards in and around Maiduguri (Figure 2), and it provided the first layer of the RLD. The list of the wards and coordinates of the centre is given in Table 9. In addition, three super wards of Gwange, Bolori and Shehuri were added to provide for the cases where only the ward is mentioned without the suffix (I or II). The OSM provided eight layers of information, namely: places, points, railways, roads, waterways, buildings, land use and natural (Table 10). Although a total of 106249 records are obtained from the OSM, only 579 have names that could be used in the reference data (Summary in Table 9). There were 29 repeated names between the layers of OSM, hence only 550 could be used in the reference.

The process of using the RLD was also a process of updating (As shown in the next section). Thus the current report of the RLD is tentative. It has 683 records, six fields (Reference Number, Type, Source, Name, latitude and longitude). Eighty percent was sourced from OSM, 5% from OSGOF and 15% from Google Earth. An Excerpt of the RLD is given in Table 13.

Table 9: Location of centroid of the Wards of Maiduguri and Surrounding

S/No	Ward	Latitude	Longitude	S/No	Ward	Latitude	Longitude
1	Ajiri	11.6507	13.22702	19	Koshebe	11.9942	13.3215
2	Alau	11.7349	13.21822	20	Lamisula	11.844	13.1337
3	Auno	11.829	13.04945	21	Limanti	11.8456	13.1637
4	Bolori I	11.8679	13.1036	22	Loskuri	11.8345	13.4172
5	Bolori II	11.8905	13.12744	23	Mafoni	11.8495	13.1491
6	Bulabulin	11.8354	13.15871	24	Maimusari	11.8386	13.2062
7	Dala	11.7426	13.05396	25	Mairi	11.805	13.2065
8	Dusuma	11.8728	13.26751	26	Maisandari	11.7963	13.1145
9	Fezzan	11.8433	13.15486	27	Mashamari	11.8508	13.1909
10	Galtimari	11.754	13.13067	28	Old Maiduguri	11.8977	13.1683
11	Gamboru	11.8568	13.16573	29	Shehuri North	11.8527	13.1596
12	Gomari	11.8269	13.07928	30	Shehuri South	11.8447	13.1587
13	Gongulong	11.9566	13.1454	31	Tamsu Ngamdua	11.8014	13.2941
14	Gwange I	11.8406	13.16937	32	Yajiwa	12.008	13.0392
15	Gwange II	11.8309	13.1671	33	Dalori	11.7638	13.2766
16	Gwange III	11.8196	13.1743	34	Gwange	11.8277	13.1712
17	Hausari	11.8394	13.15057	35	Shehuri	11.8507	13.1594
18	Khadammari	11.9451	13.24429	36	Bolori	11.8792	13.1156

Table 10: Summary of OSM layers

	Number of Records	Number with Records with name
Places	522	352
Points	276	106
Railways	25	0
Roads	13725	42
Waterways	25	7
Buildings	90790	50
Landuse	824	20
Natural	62	2
Total	106249	579

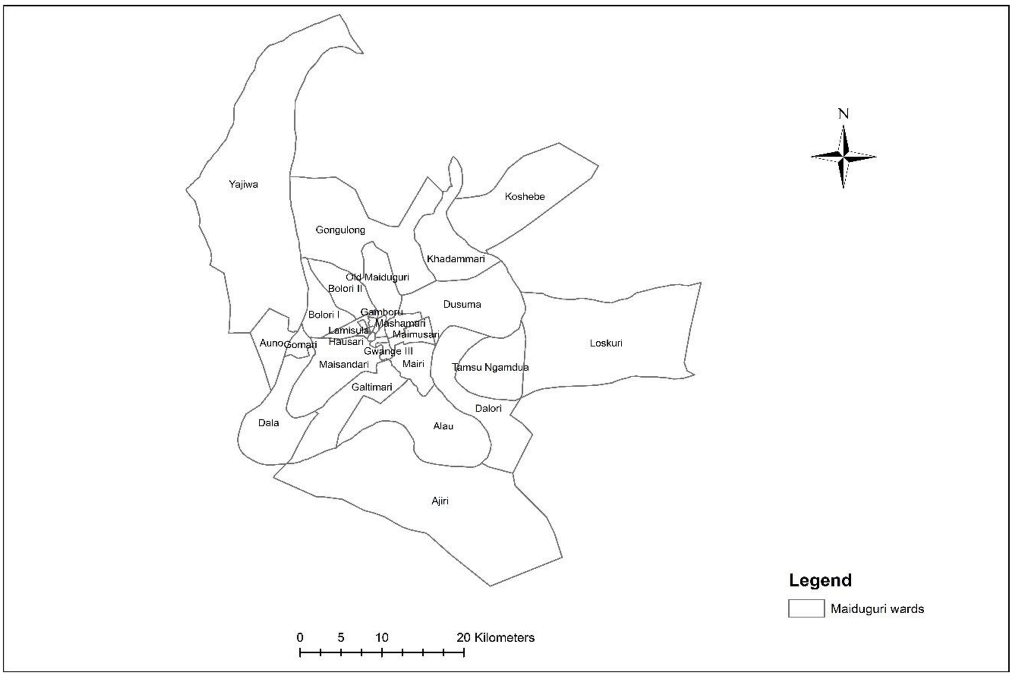

Figure 2: Wards of Maiduguri and its neighbourhood

Table 11 Excerpt of Reference Location Dataset (RLD)

	Reference Number	Type	Source	Name	latitude	longitude
662	R00594	Place	OSM	Wudimari	11.882705	13.21589
663	R00595	Place	OSM	Wujemari Kura	11.707638	13.317465
664	R00697	point	GoogleEarth	Wulari	11.849195	13.133599
665	R00696	point	GoogleEarth	Wulari Jerusalem	11.843266	13.136408
666	R00596	Place	OSM	Ya Amsari	11.837316	12.902177
667	R00639	area	OSGOF	Yajiwa	12.008048	13.039175
668	R00597	Point	OSM	Yajiwa Camp Kusheri Refugee Site	11.788707	13.130615
669	R00598	Place	OSM	Yaleri Kurnawa	11.668846	13.392501
670	R00663	point	GoogleEarth	Yan Robobi Gamboro Market	11.852029	13.17225
671	R00599	Place	OSM	Yangomari	11.688028	13.228703
672	R00600	Place	OSM	Yar Dole	11.9297	13.2428
673	R00601	Place	OSM	Yarimari	11.923255	13.224212
674	R00602	Place	OSM	Yauri	11.7648	13.2155
675	R00710	point	GoogleEarth	Yauri	11.764984	13.215495
676	R00603	farmland	OSM	yee haw land	11.870361	13.154666
677	R00604	Place	OSM	Yuramti	12.008321	12.94649
678	R00605	Place	OSM	Yuramti	11.674825	13.050723
679	R00606	Place	OSM	Yuti	11.74599	13.239892
680	R00644	point	GoogleEarth	Zabarmari	11.930011	13.240131
681	R00749	point	GoogleEarth	Zabarmari	11.930011	13.240131
682	R00647	point	GoogleEarth	Zajeri	11.870672	13.127667

2.4 Georeferencing: Matching Extracted Data with RLD

The 389, 255 and 258 records extracted from GTD, ACLED and GEDevents, respectively were searched in the RLD that was built from the OSGOF and OSM. When Excel Fuzzy Lookup was applied to the extracted GTD and RLD, 26% could not be found, 19% were wrongly assigned location, 31% were partially correct in the assignment and 24% absolutely correct. On examination of the partially correct assignment the reasons are: misspelt names such as Gombouru in the GTD and Gamboru in the RLD. There were cases where shorter version of the name was used in either the data base or the RLD, for example, Muna and Muna Dalti. Table 12 is an excerpt of the result of comparison of GTD and reference data. It should be noted that both the reference data and the sources of the extracted data are not thoroughly reviewed to have standard names.

ACLED and GEDevents showed similar characteristics, the summary of the matching is shown in Table 13. Most of the locations extracted were Matched. However, the average of 30% not matching indicated the need to improve the RLD

There were 93 unique names at the initial use of the RLD, that could not be found or were wrongly matched. The location of most of these names were determined using the Google Earth Pro. A simple search produced many of the locations, and others were determined based on the researcher's knowledge of the area, thus the process produced an updated RLD

Table 12: Execerpt of the report of the Matching of Extracted GTD data and the RLD

Extracted	Reference	Latitude	Longitude	Similarity Index	Summary %
Gwaidomari				0.0000	No result, 26%
Wahabi mosque				0.0000	
Sinimari Church of Christ				0.0000	
Bayan Quarters	Adam Kolo Bayan Makaranta Camp	11.833244	13.186372	0.5187	Wrong matching, 19%
Balgawe Ajiri	Bulabulin Baligawe	11.710507	13.12171	0.5432	
Muna Garrage	Muna Garage El Badawe Refugee Site	11.874751	13.250862	0.8387	
Kaleri Tomsu Ngamdu	Kaleri	11.68285	13.062116	0.8750	Partial Matching, 31%
Muna	Muna Dalti	11.8684	13.2445	0.8800	
Gombouru	Gamboru	11.856765	13.16573	0.8889	
Njimtilo	Njimtilo	11.8495	13.0219	1.0000	Correct Matching, 24%
Muna Dalti	Muna Dalti	11.8684	13.2445	1.0000	
Yauri	Yauri	11.7648	13.2155	1.0000	
Gamboru	Gamboru	11.856765	13.16573	1.0000	

The extracted data from GTD, ACLED and GEDEVENTs were respectively matched with the updated RLD. The summary of the matching performance is given in Table 12. Although there were high percentages of failure to match, that indicated the need for further field work other than the use of Google Earth.

Table 13 The Performance of Matching of the Dataset

Dataset	Correct Matching	Extraction records
GTD	262 (67%)	389
Gedevents	192 (74)	258
ACLED	165 (65)	255

The combined results from the three dataset were used to generate a map (Figure 3) that shows the various locations terrorist activities took place in Maiduguri.

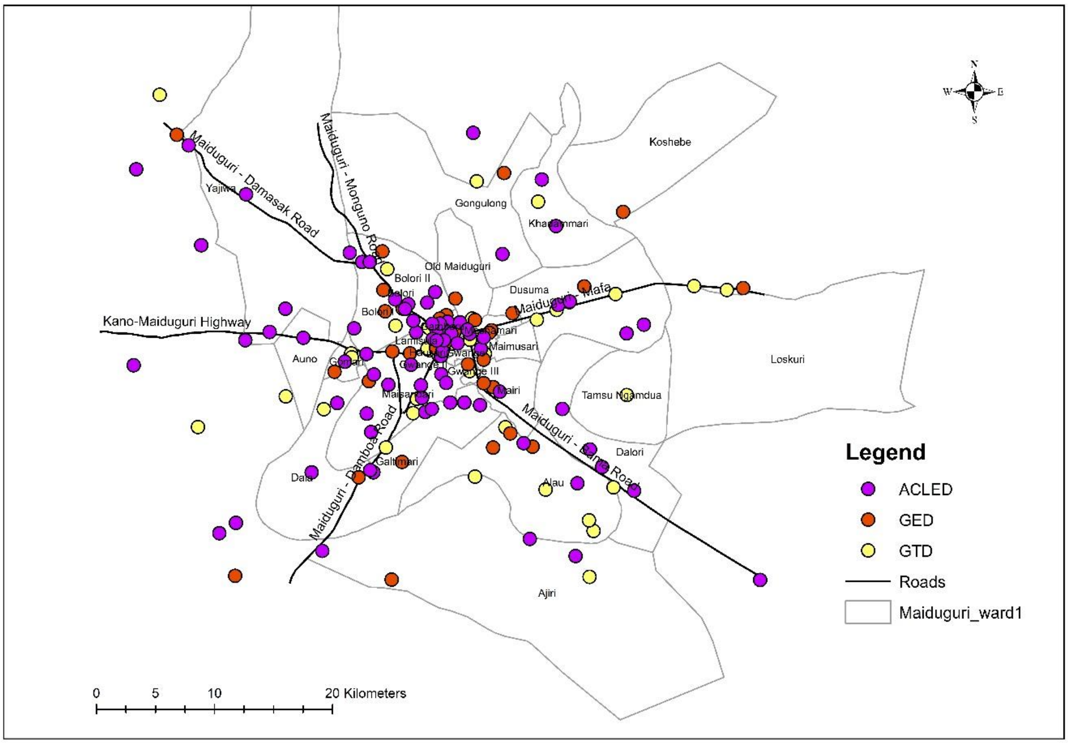

Figure 3: the location of activities of terrorism in Maiduguri

Conclusions

Textual reports are used to extract the location information of a suburban. Geospatial investigation at suburban level has become possible with the determined coordinates. Hence the geospatial analysis of insurgency in Maiduguri will be richer when the coordinates in Maiduguri are known. This work demonstrated how the geospatial precision of public data can be improved. Global terroristic attack tracking data provides a viable means suburban location of terroristic attacks from their textual reports. The preposition "in" is the most important word that can be used in extracting location of an attack. The extraction shows that a record in ACLED and Gedevents more than one name of location, and a name of a location may not be a single word, and may have different variance for same location. The process of extraction was tedious, although particular words were used to point to text that depicted location, the process does not give the location directly. The process only narrows the search, a further work had to be on selecting the name of the location of attack, which implies individual verification. This also point to need of artificial intelligence, since there are enough data for machine learning. A much easier way is to directly search already known names in a given text. That is not case here because reference data has not been developed. When such reference has been built, it will not eliminate the process of extraction but will make it secondary. The geospatial coordinates of centroid of administrative wards around Maiduguri formed the first layer of the RLD, and then OSM data were added. Since the two layers of data were inadequate, the Google earth was used to determine the coordinates of extracted location that could not be determined from layers. Building a robust RLD will shift the task of identifying using preposition to direct search for name of location. This study has provided the basis for geospatial analysis of terrorism attacks in Maiduguri: where the attack occurred over the years, and according to wards. There is need to develop local terrorism tracking system that will capture suburban location. Since the target is names of suburban, the social media in the localities are most likely to mention the local suburban names, as Hoang and Mothe, 2018 has demonstrated effort to use Tweeter. A lot of time was spent extracting the names using Microsoft Excel; it is recommended that the use of more robust Artificial Intelligence be used (Radke et al., 2019, Johansson and Nugues, 2007).

Reference

Bahgat, K. and Medina, RM. 2013. An overview of geographical perspectives and approaches in terrorism research. Perspectives on Terrorism, 7(1): 38-72

Microsoft 2022. Fuzzy lookup add-in for excel. Available at: https://www.microsoft.com/en-us/download/details.aspx?id=15011 accessed on 16/05/2022

Haklay, M. and Weber, P. 2008. Openstreetmap: User-generated street maps. IEEE Pervasive Computing, 7(4): 12-18.

Hoang, TBN. and Mothe, J. 2018. Location extraction from tweets. Information Processing and Management, 54(2): 129-144.

Johansson, R. and Nugues, P. 2007. LTH: Semantic structure extraction using nonprojective dependency trees. Proceedings of the fourth international workshop on semantic evaluations (SemEval-2007), 23-24 June, Prague, pp. 227–230.

Melo, F. and Martins, B. 2017. Automated geocoding of textual documents: A survey of current approaches. *Transactions in GIS*, *21*(1), 3-38.

OSGOF, 2022. Nigerian Administrative Map available at:

https://data.humdata.org/dataset/cod-ab-nga#, accessed on 8/3/2022

Radke, M., Das, P., Stock, K. and Jones, CB. 2019. Detecting the Geospatialness of Prepositions from Natural Language Text (Short Paper). Proceedings of the 14th International Conference on Spatial Information Theory (COSIT 2019), (Timpf, S., Schlieder, C., Kattenbeck, M., Ludwig. B. and Stewart, K. (Eds.). 9-13 September, Regensburg, Germany, Article No. 11: 11.1-11.8.

Skoumas, G., Pfoser, D., Kyrillidis, A. and Sellis, T. 2016. Location estimation using crowdsourced spatial relations. ACM Transactions on Spatial Algorithms and Systems (TSAS), 2(2): 1-23.

University of Maryland 2022. at: https://www.start.umd.edu/gtd/ accessed on 10/3/2022

Uppsala University, 2022. at https://www.pcr.uu.se/research/ucdp, accessed 10/3/2022

Waziri, M. 2012. Spatial Pattern of Maiduguri City. Researchers' Guide, Adamu Joji Publishers, Kano

Corresponding author's e-mail address: ssgarba@unimaid.edu.ng