International Journal of Sciences (ISSN 2305-3925) Nov-2012

Dilemmas in Crime Mapping: Data dearths, procedures and baseline datasets

S. Formosa
Department of Criminology, Faculty for Social Wellbeing, University of Malta, Msida

Correspondence to Saviour Formosa, Ph.D., Department of Criminology, Faculty for Social Wellbeing, University of Malta, Msida, Malta MSD2080. Email: saviour.formosa@um.edu.mt. Phone number (356)23402771.

Abstract

Creating a crime mapping system for a small state offers challenges that may defy database processing reasoning. The barriers to access, digitisation, geocoding and integration with other datasets create difficulties that require the setting up of a series of protocols to ensure data integrity. The Maltese Islands' case study delineates the processes taken up to ensure the setting up of baseline datasets in the crime, social and landuse domains. The study reviews data inputting, digitising and converting, error generation and geocoding. Data reliability, aggregation and analysis processes are covered due to their intermediate function between the main focus of the baseline data creation and the eventual research output as disseminated through online technologies. Malta's data processing has proven to be a tough slope for researchers, however the creation of the baseline data layers has ensured that new incoming data can be integrated in the systems quicker and be analysed at faster rates.

Keywords: baseline, dataset, Maltese Islands, crime mapping, geographical information systems, spatial statistics

1 Introduction

Most studies in the social sciences employ either qualitative or quantitative analysis with limited emphasis on the spatial component. The main emphasis in this study is based on a review of the quantitative methods with a strong spatial analysis employed in the creation of the first crime baseline study in the Maltese Islands. The concept of analysing crime from a spatial viewpoint allows the researcher to envisage new inputs to knowledge without sacrificing the statistical construct. However, the process to initiate such analyses requires a review of the data availability, access and integrity which would allow for an understanding crime. This study takes up the data review, processing and dissemination aspects as well as delivers some outputs emanating from the information system structuring. The paper covers the work carried out to digitize and in turn geocode Maltese crime over 50 years from 1950 to 1999.

At the initial stages of the research there was little, if any, data available on crime in Malta and definitely none in a digital format. Such a situation entailed the initiation of a long process of data mining, gathering, cleaning, data base design and geocoding prior to any attempt at analysis.

In addition, the author's decision to adopt a 100% coverage (i.e. analysis of all the available data rather than sampling) of any data gathered for the islands increased the pressure on the data gathering process. This was done to make sure that any relationships in such a micro-state as Malta (316 km.sq. 410,000 population) may be analysed to the most detailed scale, as some crime types may only occur in one area that is synonymous with a specific activity and that may fall out of any sampling methodology used. In effect, having recognised the implications of such a decision, this process was initiated in 1997, with six full years of data gathering and two for data cleaning and georeferencing, which process resulted in the generation of the first maps, which data is currently being updated with data from the 2000s.

This method was followed as it ensured that it would lay out a solid foundation for any future research on criminality in Malta. It would also ensure that enough material has been gathered in one repository to enable an enhanced analysis process, both within tabular statistical and spatio-statistical analysis.

2 Background to methods used

In the analysis of crime, Bottoms and Wiles (2001) state that one of the problems that has 'bedevilled' environmental criminology theory since the Chicago School (Shaw and McKay, 1969) and is still unsolved concerns the relationship between the various procedures used to analyse crime. They state that 'social activity' needs to be analysed in terms of a multi-level approach: 'hot spot' analysis dealing with concentrations of offences, person-person interactions dealing with offender-victim issues, and response of the social actors to the physical aspects of the location (Parnaby, 2006; Hirschfield, 2011). In effect, this brings together quite a diversity of methodological issues. The 'hotspot' analysis necessitates the activation of a quantitative approach using high-end GIS and statistical tools. Hotspot analysis can be analysed through the NNA and NNH as well as the Getis-Ord G Statistic, the Moran I Statistic, the K Function Statistic, GAM clustering methodology (2000) and Geographic Weighted Regression (2002).

The person-person interactions call for knowledge of these same interactions either through ethnographic studies and participant observation research, amongst others. In the case of the response of the physical actors the interview or survey method would be appropriate. This methodology was used by Wang (1999) in his study that included the exploratory (quantitative – forecasting at macro level) and explanatory (qualitative – daily social life that individuals anticipate).

In the Malta study, the quantitative approach is taken; the initial study investigates the offender rate over a 50-year (1950-1999) period. This is complemented by a detailed study spread over 10-years (1990-1999) focusing on both offender and offence data.

2.1 The quantitative process: barriers to access

The study relies heavily on the quantitative approach. This is a natural constraint once the decision to go for a spatial perspective was taken. The approach called for a highly numeric approach particularly since the whole Maltese parameter was employed. It also entailed gathering data from a large number of analogue sources, again rarely digital.

Though complete parameters were chosen, the process itself entailed some high level use of sampling methodology. This was based on different methods according to the type of study being undertaken, such as purposive sampling where the offenders chosen were those who had been incarcerated as against all 'guilty' offenders. The method was also employed when choosing offence data from the incarcerated dataset where only the last decade of the century was taken. This sampling method was both scientific and practical as data on all sentenced

persons is not made available and where it is available, consistency in attribute input is not always reliable.

This reliability issue crops up due to an inconsistency in data recording at the Courts of Malta, where data has yet to be centralised and quality controlled. In addition, data from the Courts is difficult to acquire access to and are dispersed in a number of court rooms, departments and sections. This said, the latter situation is being rectified where since 2001, on-line sentencing information is being disseminated in Maltese (2012). Sentences are inputted as dictated by a judge or magistrate and the documentation does not have any sociological data as needed for research purposes such as recidivism and background studies. Dissemination of such data would break the legislative safeguards on data protection and human rights, amongst others.

In view of the above restrictions, a decision was taken in 1997 to identify the available crime data sources and to choose those available for research. Once certification was made that court data was not accessible and that no comprehensive digital data existed in the Courts, that source was considered too tedious and bureaucratic to embark on and was subsequently discarded. The next data sourced included generic crimes statistics (available from COS[1] in book format with digital versions since 1998). crimes reported to the police and prison statistics available in ledger format for over 150 years. Both the police and prison data were chosen and a process of digitisation initiated.

The decision to analyse prison records was taken as it is the only crime data repository where information is held literally “secure behind bars” and in one place. This dataset provides data on the ‘hardcore’ or at least the most reliable source of persons who have been truly convicted by law and have a proven direct and severe relationship to crime. Analysing suspects and arraigned persons may depict a different outcome of what makes an offender as a large number would be acquitted. Thus those convicted have gone through the whole process and have been filtered by police, society and law (Formosa, 2007).

This also bears on the fact that studies on crime need also to analyse whether incarceration is providing enough deterrence value in the strive to contain crime. Wolpin (1978) and Entorf (2000) in their studies of European studies lists 5 variables that should act as deterrence variables, which variables can be termed to represent the different types of sentencing levels that can serve as surrogates for criminal analysis.

a) clear up rates (crimes cleared by police);

[1] Central Office of Statistics renamed to National Statistics Office, which publication series "Abstracts of Statistics” hosted Crime-related data, which data was discontinued since 1998.

b) conviction rates (proportion of arrested pleading guilty or are convicted);
c) imprisonment rates (proportion of guilty who are imprisoned);
d) recognizance rate (proportion of guilty who are on recognizance);
e) fine rate (proportion of guilty who are fined);
f) average sentence (average length of the court sentence for those imprisoned).

In the Malta case some of the data may not be available, however Entorf *et* al (2000) state that normally at most only two variables are used in most studies.

In the case of this study, Awaiting Trial offenders were not included due to the fact that a number of these persons may be acquitted or their cases remain open for a considerable period of time as well as being converted to Full Imprisonment files once convicted, a process that may result in double-counting. Thus the c) category was chosen for this study.

3 Data gathering exercise: the process and the hurdles

Inherent to an analogue-digital transfer process, the data gathering process took up the greater part of this study, particularly due to the lack of any digital data. This was particularly critical in the crime-related field such as convicted offender and convicted offence data, and national statistical compendiums. Where digital crime data was made available, a process of validation as well as geocoding and georeferencing was undertaken.

The main issue that needs to be mentioned at this stage concerns data currency. As the study looks at 50 years of crime, gathering this information at national level was possible. In other cases the data is only available for the last ten years of the century, particularly since offence data was only kept in storage since the late 1980s and quite a few of the initial files were badly deteriorated having been eaten by mice and other creatures. In other cases data is only available since 1998, such as the police reported data. However the author managed to acquire data from the 1995 Census of Population and Housing at enumeration level. This makes it possible to analyse offender-offence relationships as against the socio-economic data of the 1990s.

The next section reviews the methodological process employed in the data gathering process.

3.1 Source identification

Digital data was even harder to find and most has to be inputted from scratch. In fact all data pertaining to this study except for the 1998-2012 police data on reported crime has been inputted by the author. Data from the NSO crime statistics prior to 1990 was also manually

inputted. GIS data on crime is virtually nonexistent, except in the MEPA planning system where infringements on development (that may be taken to Court) are mapped in a GIS system, but are technically not relevant to this study (2000).

The data sourcing was one of the main concerns of this study, mainly due to access issues, analogue formats, permissions required, data-mongering and major bureaucratic issues. Whilst identifying the data sources for the different themes entailed a straightforward approach, the main problem was getting to the data.

The first step taken involved coming up with a list of datasets that would prove essential to this study and acknowledging the fact that it may not be gathered in time for the analysis. Crime data in Malta exists mainly in analogue (paper) form and to a lesser but greatly restricted extent in digital form. Malta has a history of data-mongering where each data creator is called a data guardian who does exactly that: data is hoarded and made inaccessible to any but the holder. This culture is giving way to a more open system where data availability on the web has helped break the hoarding mentality. Data is made available without compromising individual protection. Citizens are protected through the Data Protection Act (Chapter 440) enacted in 2001 and brought fully into force on the 15th July 2003 (OJ, 1995; Government of Malta, 2001). The data is administered by persons called *"controller of personal data who alone or jointly with others determines the purposes and means of the processing of personal data"*.

Social data was sourced from a number of organisations including the National Statistics Office, the Ministry for Social and Family Affairs, the Education Department, the Health Department and Planning Authority. Due to the sensitivity of these datasets, access was requested directly from the Ministers involved or Director Generals as well as the Prime Minister's Office.

Land Use data was sourced either directly from the organisations archiving them or indirectly through value added datasets acquired by third-party organisation from the original sources. The latter were primarily sought from the Planning Authority – later called the Malta Environment & Planning Authority, Water Services Corporation, Local Councils Department, National Statistics Office, and Malta Transport Authority.

Permission was sought and approved quite rapidly by the Director of the Planning Authority. The tabular and spatial data entailed cleaning, organising, aggregating and restructuring. This was necessary as the data was not originally intended to provide for social matters, effectively rendering most of the datasets 'ineffective' or unusable. However, new maps were created based on the datasets, with major recoding in such layers as address point database, street layer and enumeration

areas.

When crime data is sought, however, analogue data was and is still very difficult to access except for National Statistics Office publications and parliamentary questions results. The Police Annual reports are not made available to the public and have not been made available even for this study. Data on court cases follows the same route and only since 2001 have cases started being published through a web-enabled database portal though in some instances records are incomplete since not all courts insert data in the system. To make it even harder, all documentation and information pertaining to offenders has a timestamp on it and information on persons who were/are incarcerated have an 80-year moratorium on them thus making the data for this study practically impossible to access. The data for the offenders of the 1950s will only be available in 2030 whilst the data for the 1990s which is needed for the offender-offence analysis will be available next in 2070. Special permission to access and digitise this restricted data was sought and granted from a series of Justice Ministries and CCF Directorships.

Finally, crime data was sourced from the National Statistics Office, Corradino Correctional Facilities (CCF - prison). Ministry for Justice and Home Affairs, Education Department, Police Force, and the National Archives. Once again, the sensitivity of the data called for highly specific access requirements such as access to prisons, police headquarters and other sensitive individual data holders. In return, the results from the crime classification exercise, as created by the author, were forwarded to the Police IT department for consideration in their PIRS (Police Incident Reporting System) categorisation. The Classification system was built on the initial Maltese category pre-1997 and post 1998 as well as from the 2007 changes to PIRS lists as well as the UK categories as extracted from the Huddersfield Kirklees 4th Crime Audit (2001).

In the crime data case, other data gathering problems faced by the author were policy and legislation related such as access to data, access to prisons, access to personal data, and the normal bureaucratic issues such as that faced in the concerned government departments, prison and crime-related entities. The process was both enabled through contacts and goodwill by individuals within each organisation and on the other hand impaired by other individuals who either sought to reduce access or in one case to terminate the data gathering process, quoting prison regulations. The latter situation was diffused through a meeting with the Minister for Justice and Home Affairs who re-granted access to the prisons. Other not-so-drastic bureaucratic issues but equally disrupting were re-access to prison data once a change in government occurred: new ministers, new directors of prison, and other new officials.

Other data was gathered through surveys and questionnaires. One such was termed the Dark

Figure of Crime Survey. The survey was distributed to 300 MEPA employees who were asked to report any crimes over 5 years and whether they filed reports to Police. However, both the sample and the reply rate was too low to enable reliable analysis and the author decided that it would not be included, though the framework is now ready for a larger run post this-study.

The above exercises sought and reached the guidelines that sought to review the data availability in the Maltese Islands in the fields of crime, social issues and land-use aspects. It sought to review and compile the data into digital format for future studies and analysis. This was possible through the conversion from analogue to digital format. As expected, the digitisation process took up the larger part of the research period.

For want of a better data source…

Crime data from newspapers was not availed of as generally short reports are given and mainly sensational, shocking and serious cases are given precedence such as homicides, rapes, hold-ups and molestation.

3.2 Data inputting, digitising and converting

The data inputting process was a tedious one, requiring various operational logistics. This process took place between 1997 and 2005 with new data being currently inputted up to 2012. Data gathering from the convicted prisons necessitated the use of manually inputting the individual offender and offence attribute data into a spreadsheet, each time necessitating checks to bring out the data from the prisons as a laptop was not allowed within the organisation. In the case of the National Archives, the process involved manually copying the ledgers and digitally inputting them in the spreadsheet once at the office. Also, data from the National Statistics Office (national crime data) was inputted as OCR technology is still not 100% reliable in the case of converting tabular data.

Where data was made available in paper map format such as a bus routes maps, this was manually digitised and saved directly in spatial format. Other data, where possible, was scanned, georeferenced and digitised.

Where digital data was made available, this was converted to the appropriate format and eventually to spatial format. In cases this necessitated conversion from textual to tabular to spatial formats in the process' full entirety. The textual to tabular conversion is necessitated to remove artefacts and align the data, whilst the tabular to spatial data is less strenuous except where geocoding

issues are concerned as explained below.

3.3 Data cleaning and verification

Next to data inputting, data cleaning and verification is also a tedious phase of any research of this scope, where large datasets are involved. Whilst the manually inputted data were double checked during the inputting phase and validated through a series of SQL queries, the data sourced in digital formats necessitated another approach. This data was individually checked for inconsistencies, the major issue being street names errors (vital for geocoding purposes). missing data, and other inconsistencies.

This is a slow aspect of the project, particularly due to the numerous multiple-interpretation street names errors, alphanumeric errors in location names, offence information, as well as intentional errors by data-inputters[2]. This process entailed going through the whole list that did not match the official street names and interpret those same names and locations.

This process was repeated for the welfare data where correctness is even more vital due to the fact that persons receive benefits on a monthly basis and incorrect addresses mean non-issuing of cheques vital to such persons. In this case the issue was mainly identifying the local names of streets, multiple street names, alley-correction (integration of alley names with connecting street names) and other quality assurance measures.

In the case of available spatial layers, data cleaning was also necessitated due to the need to identify lacunae, overshoots, gaps, intersecting and overlaid polygons as well as missing data. This is important especially where Census data is involved due to the intersecting street centrelines and their subsequent polygon structure having a considerable part overlapping the corresponding street polygon. The resultant data can be misleading and depict an erroneous output leading to a wrong interpretation.

3.4 Geocoding procedures

Once the data was digitised, the next phase entailed the processing of all data to spatial format. This is possible through a process called geocoding where data in tabular format is given a spatial construct by assigning each data item some kind of spatial element. In most cases this

[2] These included genuine typos (such as Valletta written in over 25 versions, for example Valetta, Valleta, etc) to near illiterate interpretations (such as Hookham Frere Street being written as Hook Ham Frier Street) to the downright witty and mischievous (such as an unknown tomb in the main cemetery written as Death Row, Addolorata Cemetery). Though eliciting a laugh the number of such errors meant time-wastage and necessitated individual record checking, reaching up to 99,575 for the Police reported crime dataset, with an average error adjustment in 40% of the streets.

entailed assigning a point element such as a street centroid (central node) or assigned to a polygon such as an Enumeration Area or Local Council.

This process is once again time-consuming and highly challenging, especially since no readily available software existed that could cater for the specific requirements of this study. Malta has no street-gazetteer, which is a data layer that would include street segments and dwelling numbers. The author sought to geocode a number of points along a street centreline when only the street address (with no address location given) was available. No *mbx* programme existed that automatically disperses a number of points randomly around a centerline, each having the same locational attribute (such as a street name). This induced the following process and rationale:

a) creating a buffer around the street centrelines in order to prepare a series of polygons in which a number of points can be generated, which points would serve as anchors for the addresses (Figure 1);
b) converting the buffers from a vector polygon to a raster grid file, which process actually creates a base of pixels for later conversion to the points described in a);
c) converting the grid file to a series of points (Figure 2);
d) processes a to c would have retained the attribute data, thus the next step is to clean and check the attribute generated in the raster-to-vector conversion. This was accomplished through a join with the centreline buffers through manual or automatic means (Figure 3);
e) once each point was validated, a unique numbering system was created so that each point could be queried individually;
f) cleaning the non-geocoded data (Figure 4);
g) as not all street names were captured through this method, other technologies were used. These included on-line searches such as checking for missing data in MEPA's mapserver (MEPA, 2000, MEPA, 2012) and other on-line datasets (Figure 5);
h) once the street-name validation was concluded, the next step sought to link the non- spatial data layers such as the police reported crime dataset to the unique points generated as per above process. This entailed the geocoding of the non-mapped layer into the dispersed points (refer to c) through a simple geocoding exercise following the insertion of the unique numbering system in the same dataset in line with e) (Figure 6);
i) Overlaying the geocoded data onto other base data for verification purposes (Figure 7).

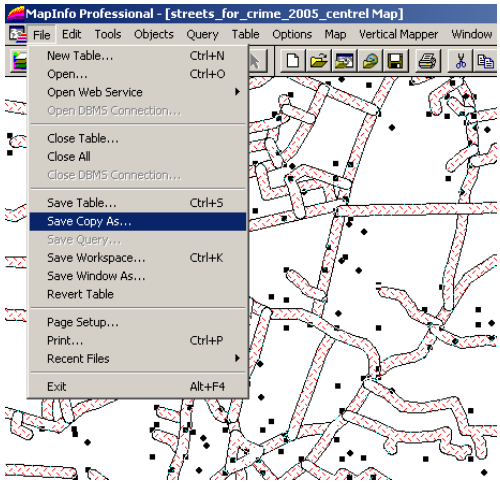

Figure 1: Creating a buffer of 5metres for each street helped to develop a series of raster and vector maps that generated points for geocoding anchors

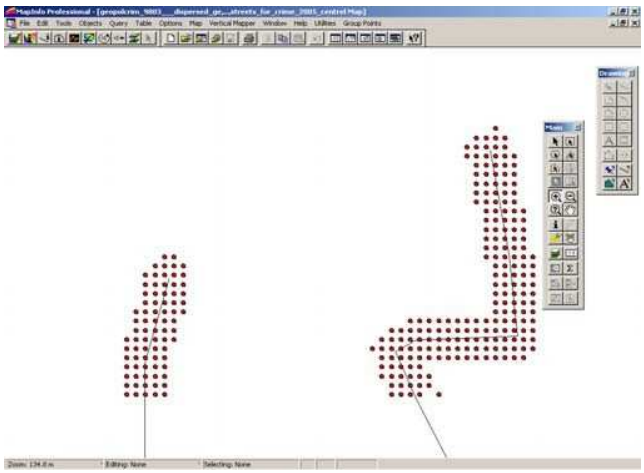

Figure 2: Rasterising the buffer points

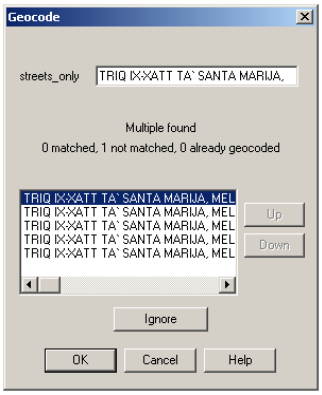

Figure 3: Checking attribute data for matching to centreline through manual or automatic means

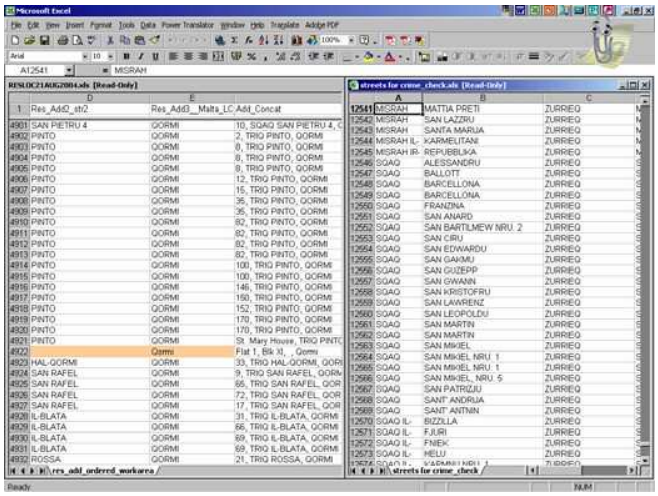

Figure 4: Cleaning the tabular data prior to geocoding

Figure 5: Checking for missing data in MEPA mapserver and on-line datasets

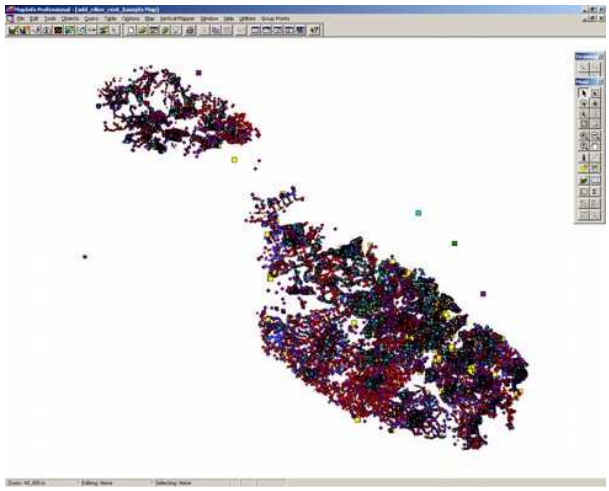

Figure 6: Geocoding the dataset based on dispersed points within the buffer

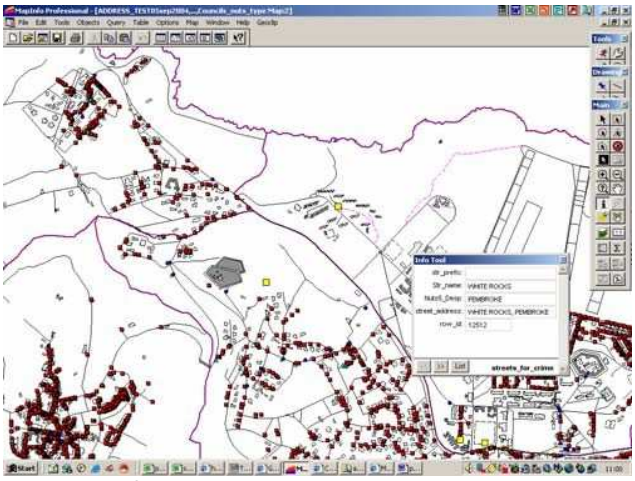

Figure 7: The final Result overlaid on other base data layers

A lineage depicting the process undertaken to geocode a map was created, where a manual for the creation of such datasets was drafted. This process sought to investigate the possibility of preparing the base data for crime analysis in spatial format, through the process of geocoding and georeferencing crime, social and landuse data into spatial format and creating the first Maltese crime, social and landuse-related base-maps.

3.5 Data available for the study

The generic data categories were made available to the researcher as listed below:

- Offender data
- Offence data – related to CCF data
- Offence data - Police data (reported crimes)
- Education data (absenteeism)
- MEPA (development data)
- NSO (Census and Abstract of Statistics) data
- Welfare Data
- Landuse Data

The full list of datasets employed within this study is listed in Table 1. All the datasets have been converted to spatial format except for the Abstract of Statistics data and the Household Travel Survey that was used for reference only and was not needed in the integration process for further analysis.

Table 1 lists all the datasets used, their time-period, the most detailed available spatial level and format. The crime data spans 50 years, the landuse data mainly relates to the late 1990s and early

2000s, the social data is mainly related to the 1990s with some 2000s data such as the welfare benefits dataset.

Table 1: List of Datasets created through the research process

Dataset Topic	Time Period (Start - End)	Resolution	Format
Crime			
Abstract of Statistics	1950 - 1999	National	Tabular
Convicted crime - offenders	1950 - 1999	Address	Spatial
Convicted crime - offences	1990 - 1999	Address	Spatial
Education - non-attendance court reports	2000 - 2001	Address	Spatial
Reported crimes to Police	1998 - 2003	Address	Spatial
Police Stations	2002	Address	Spatial

Dataset Topic	Time Period (Start - End)	Resolution	Format
Socio-economic			
Cemeteries	2003	polygons	Spatial
Clinics & Hospitals	2000	Address	Spatial
Electoral Register	2002	street	Spatial
Housing Estates and HOS	1980 - 1993	polygons	Spatial
Household Travel Survey	1999		Tabular
Libraries	2000	Address	Spatial
Schools	2000	Address	Spatial
Sports in Schools	2000	Address	Spatial
Census Population	1995	EAs	Spatial
Census household income	1995	EAs	Spatial
Census shared dwellings	1995	EAs	Spatial
Census ownership dwelling	1995	EAs	Spatial
Census persons per room	1995	EAs	Spatial
Census amenities	1995	EAs	Spatial
Census vehicles	1995	EAs	Spatial
Census children per household	1995	EAs	Spatial
Census citizenship	1995	EAs	Spatial
Census literacy	1995	EAs	Spatial
Census educational attainment	1995	EAs	Spatial
Census educational qualification	1995	EAs	Spatial
Census household unemployed children	1995	EAs	Spatial
Census economic activity by sex	1995	EAs	Spatial
Census occupation	1995	EAs	Spatial
Welfare Benefits	1995 - 2003	street	Spatial

Dataset Topic	Time Period (Start - End)	Resolution	Format
Landuse			
Basemap features	1988 - 2005	Point	Spatial
Address Point Database	1995	Address	Spatial
Bus Routes	2003	Buffer Polygons	Spatial
Road capacities	2000	street	Spatial
Development constraints - protected zones	2003	point and polygon	Spatial
Local Councils	2003	NUTS 5	Spatial
Development Zone	2000	point and polygon	Spatial
Enumeration Areas	1995	EAs	Spatial
Islands	1995 - 2005	NUTS 3	Spatial
Local Plan Areas	1990	NUTS 4 - equivalent	Spatial
Place Names	2003	NUTS 5	Spatial
Planning Applications	1993 - 2005	Address point and polygons	Spatial
Industrial Areas	2001	Address point and polygons	Spatial
Eateries	1999	Address	Spatial
Maltamal - dwelling vacancy and population	1996	street	Spatial
Planning area footprint	1993	point and polygon	Spatial
Retail units	1999	Address	Spatial
Road Inventory: includes street furniture	1993 - 1995	point	Spatial
Road system	2000	street	Spatial
Supermarkets	2001	Address	Spatial
Town Centres	1999	polygons	Spatial
Urban Conservation Areas and Village Cores	1994 - 2004	polygons	Spatial

3.6 Data Reliability

The main issue at stake that needs to be highlighted at this stage concerns the data currency that is employed throughout the study. As this is a first study of crime at high-detail in the Maltese Islands the source data is spread in many places, rarely accessible and disjointed. Every effort was taken to acquire as much data as possible pertaining to the period under study, though this was a major problem as little consistent data is available.

Harries (1974) identifies problems as regards to spatial analysis due to the fact that substantial spatial distortions can be found when analysis was carried out based on decennial censuses. If one uses the population of an area based on the census figures and then analyses crime for one year between the census periods using that same decennial data, the results would be erroneous - using 2003 crime data and compare it to the census population of 2005 and then to that of 2011. One would prefer to have the same years' data coverage, however this is only possible in an ideal world once all required datasets are acquired, converted to tabular format and

subsequently geocoded. Only then can they be maintained and an annual update be created. That phase would then allow exact period analysis across all disciplines.

3.7 Aggregation levels of data available

Spatial data layers are aggregated in a number of levels that are primarily internationally recognised (such as NUTS nomenclature) or local boundaries specified by each entity requiring such an exercise (EC, 2003) . The main aggregation levels available in the Maltese Islands are listed in Table 2.

Table 2: Spatial data Aggregations that have been used to analyse the research data

Spatial Aggregations		
Data Aggregate	**Spatial Data Model**	**Entities**
NUTS 1_2 (National)	Area	1
NUTS 3 (Islands (Malta and Gozo-Comino))	Area	2
NUTS 4 (Districts)	Area	6
NUTS 5 (Local Councils)	Area	68
EAS (Enumeration Areas – 150-180 households)	Area	1,157
LPA (Local Plan Areas)	Area	7
Police Regions	Area	2
Police Districts	Area	11
Police Divisions	Area	20
Health Centre Regions	Area	8
Rural (Out of Development Zone)	Area	3
Village Core / UCA (Urban conservation Areas)	Area	94
Dev Zone (Limits to Development Zone)	Area	129
Streets Combined	Line	7,565
Streets	Line	13,098
Address Points	Point	199,694

Very few organisations follow NUTS districting, having predated it. Usage of the aggregate layers from different entities posed new problems in analysis: in fact it seems that each organisation made it a point to be different, using different aggregations based on operational logic. This said, the police are looking at using the NUTS nomenclature as the basis for districting. This issue was solved through the generation of data at the highest detail possible: at street level. This allows any researcher to build up rather than using the top-down approach that hinders analysis once an aggregate cannot be split up further into its component parts due to

lack of data at that high-detail level. Thus, from basic street data, the different organisational districts (such as police or health) can be catered for.

At the social level, data on welfare was provided at street level, and Census data at Enumeration Areas (EAs) level, thus this ensured that the real highest-detail level of analysis is the EAs. This implies aggregating data one level up from address-point or streets to EAs level.

4 Data analysis

Data analysis was carried out through the employment of a variety of these methods, particularly Moran's I, hot spot and interpolation. It also bases its procedures on the following methods.

i) One of the best methods of analysing crime patterns is to use clustering methodology. Due to the large number of crimes occurring in a particular area, analysis may concentrate on the aggregation of these data into specific areas rather than spread them all over the town. Clustering helps in identifying areas that are hotspots for specific crime types (Felson, 2006);

ii) Another method that can be employed is the Nearest Neighbour Analysis (NNA) which helps to aggregate data based on the proximity of a crime to the nearest location of another crime (Craglia *et al*, 2000). If a crime occurs within a specific parameter of say '20m' from that being analysed, then these two crimes are aggregated, before searching for other crime/s within the next specific boundary. Once there are no crimes left within the recurrent buffers then the hotspot intensity dies out and stops. Where a large number of crimes occur in a small area the hotspot is very pronounced and cluster densities can be calculated. Figure 8 depicts an example of such an NNA interpolation based on non-serious offences in Malta between 1998-2003 transposed in 3D. The shape of the Maltese Islands is easily discernible, particularly the conurbation area. High offence counts are depicted as with red peaks in the main leisure and recreation areas and very few if anything in the rural and rural-urban boundary areas (blue and white respectively). The same methodology can be used to elicit statistical results as well as for visualization purposes.

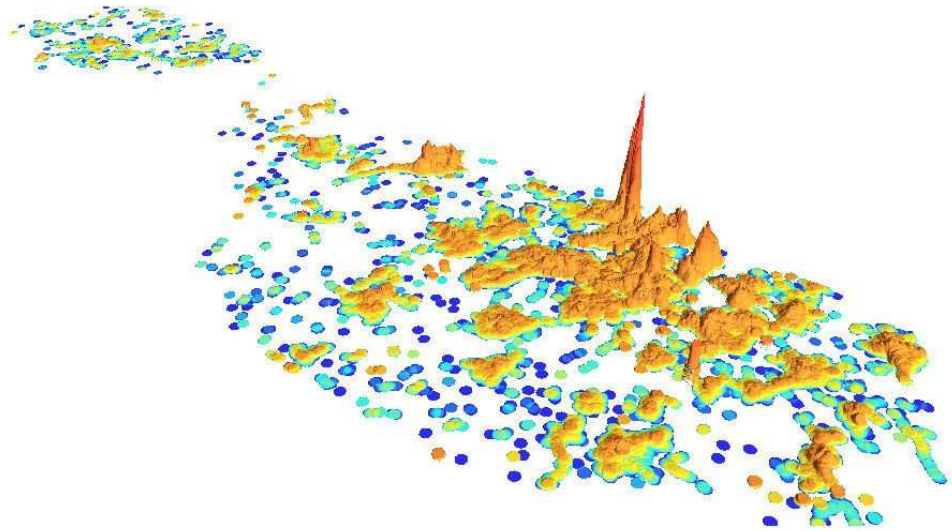

Figure 8: Interpolation of Non-Serious Offences – 1998-2003
Source: The Author

The limitations of the methodologies used such as the NNH include differing hotspot locations for different spatial aggregations employed, such as a minimal 25-point hotspot cutoff, which signifies where an ellipse boundary should be drawn once no more points falling within those thresholds are encountered. Consistency in the results is ensured as the analysis in this study employ the same threshold limits. Another limitation relates to the issue of cross-comparison of two data- layers that may have widely-differing counts, such as a 10,000 point offender data layer and a 1,000 point poverty layer. Using the same standard-deviation levels and thresholds, error generation can be reduced to a minimum.

Some ellipsoids might show areas that have high concentrations of incidences when the base data might show few data points, which result is mainly due to a multiplicity of overlapping points found within the base data layer and weighted for in the ellipsoid. Knowledge of the base data layer is required in order to interpret the results of such methodologies.

The issue of boundaries designation is highly sensitive in that the decision taken on which area level for analysis can result in arbitrary effects particularly since the same spatial entities under study may create ambiguous results where the bounds are not clear or where one could have ambiguous transition zones as well as generating some boundaries that ignore the interaction between the activities within that boundary and those external to it. Such results may

give an indication that only those areas within the resultant boundaries show incidences of clustering. This issue can be solved through the employment of sensitivity analysis strategies such as those posited by the Modifiable Areal Unit Problem (MAUP).

In addition, there are other methodologies that may have given differing results should they have been employed in this study. These include the alternative measures of hotspot analysis for example Kernel Density Estimate, Getis-Ord GI* and also multiple regression as an alternative to bi-variate analysis.

a) Kernel Density Estimate can be employed to estimate the probability density function of a random variable such as offence location. Employing kernel density estimation on a sample population helps researchers to extrapolate the data to the entire population (Wasserman, 2005).
b) The Getis-Ord GI* statistic is based on an analysis of weighted data points with an output based on a z score for each feature, employing either Euclidean or Manhattan distance. Getis-Ord GI* high z scores indicate its neighbors have high attribute values, and vice versa. The higher (or lower) the z score, the stronger the association.
c) Multiple-regression could also have been used in this research once the uni- variate and bi-variate analysis were completed, however this was not taken up for reasons for time and space constraints impinging on the research. Multi- variate analysis helps one to examine more than one predictor of an identified variable. It is highly relevant to social sciences due to its inclusion of a number of predictor variables which enable higher levels of prediction of the outcome variable.

These methods of analysis can also help researchers to use geodemographics and other social data to analyse crime and the potential target areas (Ekblom, 2008; Paynich and Hill, 2010). Use of transport and retail GIS can help to identify areas that have a high concentration of retail and travel activity that would attract 'predators'. In this scenario, GIS can help to identify the high traffic-volume streets together with the potential areas where retail is highly concentrated and would serve as an attractor for criminal elements who in turn could be studied for their journey to crime (Deale, 2011). Together with other data such as Census data, one can predict users of a service and the potentiality for crime to occur (Cairns, 1998). however, care must be taken to ensure that GIS does not end up as the end rather than a means to an end for analysis (Pease, 2001).

5 Outputs

Having laid the groundwork for the establishment of a spatial baseline dataset, the crime information system which integrates all the different datasets could be employed to help run the diverse queries across the different domains. Though not the main target of this paper, in effect, the study brings spatial statistics into social research where statistics is not yet considered the mainstay for scientific analysis in the local scene, and the spatial dimension may yet prove the jolting kick to make researchers aware of the potential of visual statistics even for other domain studies such as development planning enforcement analysis (Figure 9). The study creates its own crime-specific spatial maps, crime-landuse analytical maps as well as crime-social analytical maps and statistics based on a GIS analysis. Another output concerned the development of a map of poverty in Malta at a detailed level (EAs) based on Census data that is integrated in the model to form a part of the crime-socio-economic analysis (Figure 10). The analytical process can eventually help develop a tool that enables crime-social analysis through the integration of the relevant datasets within an integrated information system. A final output was concerned with dissemination was the Malta Crime Web-Map (Figure 11).

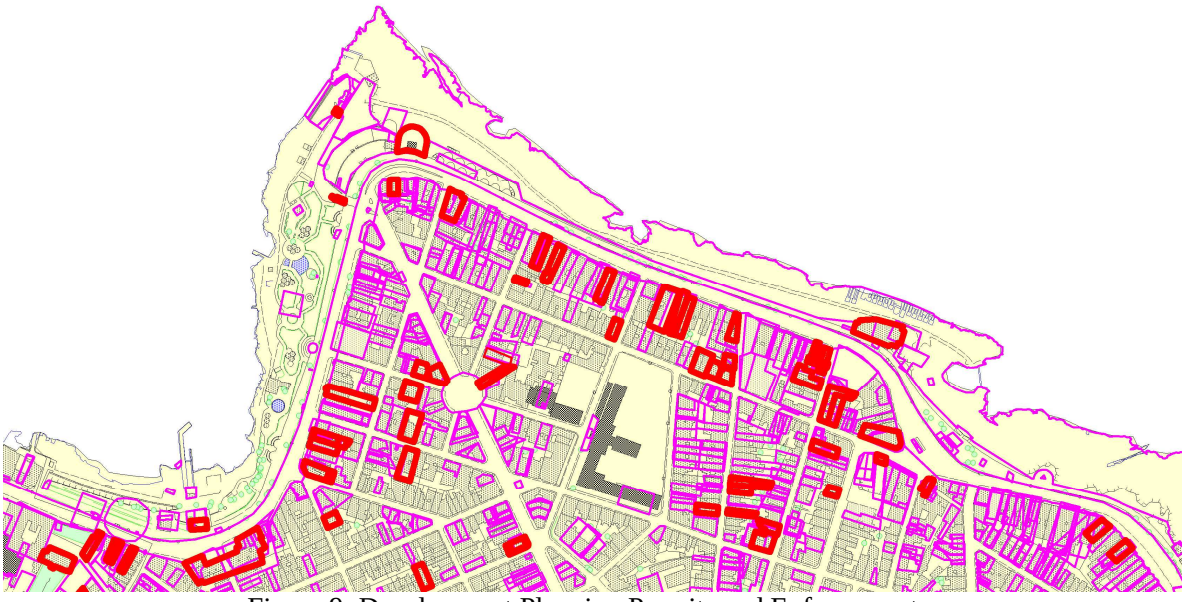

Figure 9: Development Planning Permits and Enforcements

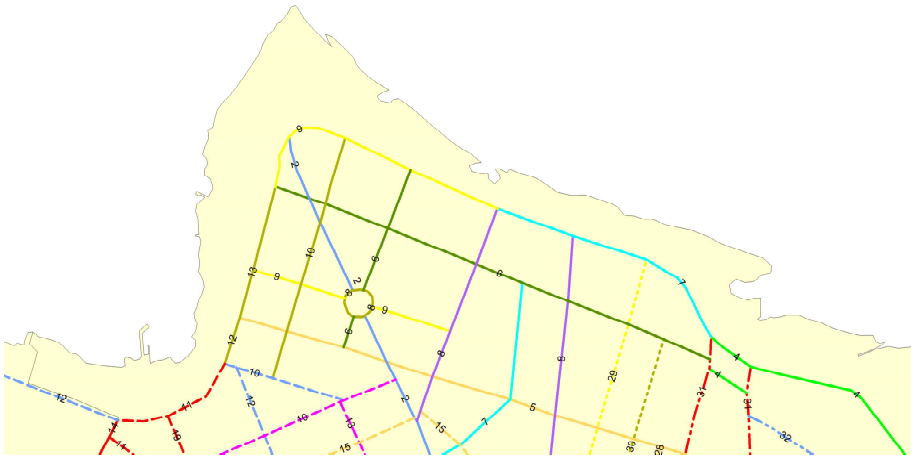

Figure 10: Census Enumeration Areas

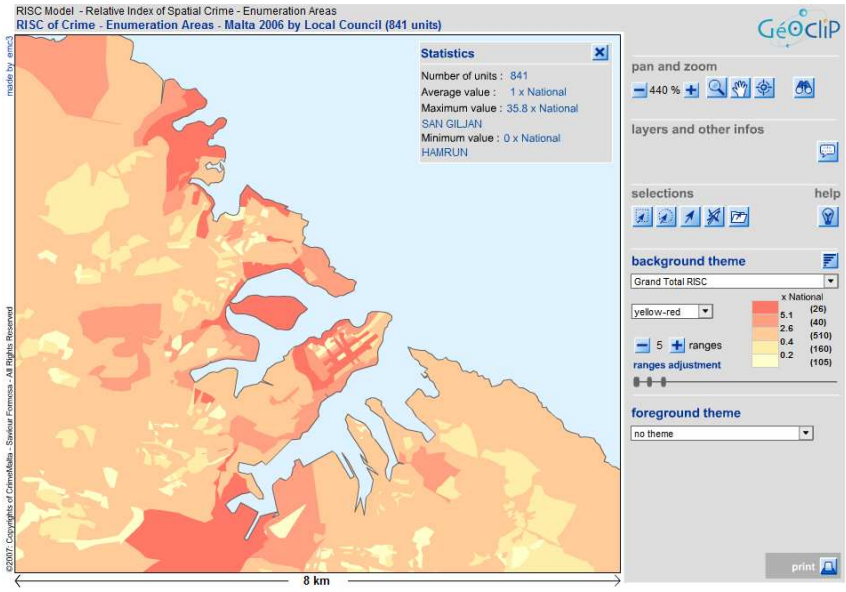

Figure 11: Crime WebMap

Source: CrimeMalta, 2012

The methodology used in this study relies heavily on the quantitative aspect, particularly due to its stress on the spatial factor. The data design process highlighted the dearth of data available and the long-winded process it entailed to produce a spatial construct for this research study. Though, highly repetitive and lengthy, the process delivered the first digital datasets and in turn the first crime maps.

In conclusion, the review of the Malta crime digitisation case study shows that is strewn with barriers

to access and digitisation error reduction requirements. There is a need to ensure that the data delivered is in line with the requirements that ensure an analysis of the interaction between all three themes: crime, social and landuse. Crime mapping renders the creation of a baseline dataset essential as it serves as the foundation for future geocoding programmes, radically reducing the geocoding time essential to the conversion of the source datasets sought from the various agencies. The development of online dissemination technologies and the respective legislative instruments have helped to ensure that the general public can view the security data from online sources without the need for highend technology.

References

Applied Criminology Group, (July 2001). Kirklees Crime and Disorder Audit - Final Report. University of Huddersfield

Bottoms A.E., and Wiles P.W. (2001). Explanations of Crime and Space, IN Evans D.J., Fyfe N.R., and Herbert D.T., (2001). Crime, Policing and Place: Essays in Environmental Criminology. Routledge, London, ISBN 0-415-04990-3

Cairns S. (1998). Promises and Problems: using GIS to analyse shopping travel. Journal of Transport Geography, Vol. 6 No. 4, pp. 273-284

Craglia M., Haining R., and Wiles P. (April 2000). A Comparative Evaluation of Approaches to Urban Crime Pattern Analysis, Urban Studies. Vol. 37, No. 4, 711-729, 2000, University of Glasgow

CrimeMalta, (2012). Crime Maps, Retrieved from http://www.crimemalta.com/map.htm

Deale, S.V., and Beken, T.V. (2011). Outbound offending: The journey to crime and crime sprees. Journal of Environmental Psychology 31 (2011) 70-78

Ekblom, P. (2008). The 5Is framework: a practical tool for transfer and sharing of crime prevention and community safety knowledge. University of Arts London Central Saint Martins. Design Against Crime Research Centre

Entorf H., and Spengler H., (2000). Socio-economic and demographic factors of crime in Germany: Evidence from Panel Data of the German States. International Review of Law and Economics 20 (2000). 75-106

European Commission, (2003). Regulation (EC) No 1059/2003 of the European Parliament and of the Council of 26 May 2003 on the establishment of a common classification of territorial units for statistics (NUTS). (OJ L 154, 21.6.2003, p. 1)

Felson, M. (2006). Crime and Nature. Sage, Thousand Oaks. In Weisburd, D, Morris, N.A., and Groff, E.R. Nancy (2009) Hot spots of Juvenile Crime: A longitudinal Study of Arrest Incidents at Street Segments in Seattle, Washington". 443 Springer Science and Business Media, LLC 2009. doi: 10.1007/s10940-009-9075-9.

Formosa S. (2007). Spatial analysis of temporal criminality evolution: an environmental criminology study of crime in the Maltese Islands, PhD Thesis, University of Huddersfield, United Kingdom.

Fotheringham, A.S., Brunsdon, C., and Charlton M. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships, John Wiley & Son

Government of Malta, (2001). Chapter 440 Data Protection Act, ACT XXVI of 2001, Valletta, Malta

Harries K.D., (1974). The Geography of Crime and Justice. McGraw-Hill Book Company, New York

Hirschfield, A. (2011). Analysis for Intervention. Chapter 22. In Tilley, N. (Ed.). Handbook of Crime Prevention and Community Safety (pp. 629-673). New York: Routledge.

Justice Services, (2012). Judgements Online, Malta Government, Retrieved on 28 October 2012, http://www.justiceservices.gov.mt/courtservices/Judgements/default.aspx

Malta Environment & Planning Authority, (2000). MEPA mapserver, Available online: http://www.mepa.org.mt/mepa-mapserver (accessed on 10 March 2012)

Malta Environment & Planning Authority, (2012). MEPA mapserver update, Available online: http://www.mepa.org.mt/mepa-mapserver (accessed on 29 October 2012)

Official Journal of the European Union, (24 October 1995): Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, L 281 , 23/11/1995

Parnaby, P.F. (2006). Crime Prevention Through Environmental Design: Discourses of Risk, Social Control and a Neo-liberal Context. Canadian Journal of Criminology and Criminal Justice, 48, 1-29.

Paynich, R., and Hill, B. (2010). Fundamentals of Crime Mapping. Massachusettes: Jones and Bartlett Publishers

Pease K. (2001). What to Do About It? Lets turn off our minds and GIS. In Hirschfield A., and Bowers K. (eds). (2001). Mapping and Analysing Crime Data: Lessons from Research and Practice. Taylor and Francis, London, ISBN: 0-748-40922-X

Shaw, C.R., and McKay, H.D. (1969). Juvenile Delinquency and Urban Areas. Chicago, IL: University of Chicago Press.Wang Y.R. (1999). Political Change and Public Security – The Prospect of Taiwan. Futures 31, 57-72

Wasserman, L. (2005). All of Statistics: A Concise Course in Statistical Inference, Springer Texts in Statistics

Wolpin K.I. (1978). An Economic Analysis of Crime and Punishment in England and Wales 1894-1967. J. Political Econ 86, 815-840 (cf IN Entorf H., and Spengler H. (2000). Socio- economic and demographic factors of crime in Germany: Evidence from Panel Data of the German States, International Review of Law and Economics 20 (2000). 75-106)