Development of Spatial Data Infrastructure based on Free Data Integration

Dr. Maythm al-Bakri
Lecturer
Department of Surveying
College of Engineering
University of Baghdad
Email: m.m.m.s.albakri@gmail.com

ABSTRACT

In recent years, the performance of Spatial Data Infrastructures for governments and companies is a task that has gained ample attention. Different categories of geospatial data such as digital maps, coordinates, web maps, aerial and satellite images, etc., are required to realize the geospatial data components of Spatial Data Infrastructures. In general, there are two distinct types of geospatial data sources exist over the Internet: formal and informal data sources. Despite the growth of informal geospatial data sources, the integration between different free sources is not being achieved effectively. The adoption of this task can be considered the main advantage of this research. This article addresses the research question of how the integration of free geospatial data can be beneficial within domains such as Spatial Data Infrastructures. This was carried out by suggesting a common methodology that uses road networks information such as lengths, centeroids, start and end points, number of nodes and directions to integrate free and open source geospatial datasets. The methodology has been proposed for a particular case study: the use of geospatial data from OpenStreetMap and Google Earth datasets as examples of free data sources. The results revealed possible matching between the roads of OpenStreetMap and Google Earth datasets to serve the development of Spatial Data Infrastructures.

Key words: SDI; OSM; Google Earth; Geospatial Data Integration; Data Quality

تطوير البنى التحتية للمعلومات المكانية بتكامل البيانات المجانية

م.د. ميثم مطشر شرقي
قسم هندسة المساحة
كلية الهندسة / جامعة بغداد

الخلاصة

لقد نالَ تطور منظومة البنى التحتية للمعلومات المكانية (SDIs) في السنوات الأخيرة اهتماماً كبيراً من قبل الحكومات والشركات ذات العلاقة. هنالك أنواع وفئات مختلفة من المعلومات المكانية مثل الخرائط الرقمية ، الاحداثيات ، الخرائط على الشبكة العنكبوتية ، الصور الجوية والفضائية يجب توفرها من اجل تحقيق المقومات الاساسية لمنظومة البنى الاساسية للمعلومات المكانية (SDI). بصورة عامة هناك مصدران اساسيان للمعلومات المكانية على الشبكة العنكبوتية: المصادر الرسمية والغير الرسمية والتي عادة ماتكون مجانية. على الرغم من النمو السريع للبيانات المكانية المجانية على الشبكة العنكبوتية ، الا ان تكامل وادماج هذه البيانات مع بعضها البعض للاستفادة منها في تطبيقات مختلفة كبناء منظومة SDI لم يتحقق بشكل فعال لحد الان. إن تبني هذه المهمة يمكن أن تعتبر الميزة الرئيسة لهذه الدراسة. يتضمن هذا البحث الاجابة عن السؤال في كيفية دمج البيانات الجغرافية المكانية الحرة وتوظيفها بحيث تكون مفيدة في مجالات مختلفة مثل تطوير (SDI).

لتنفيذ هذه المهمة تم اقتراح منهجية عامة تستخدم المعلومات الهندسية الخاصة بشبكات الطرق مثل الاطوال ، النقاط الوسطية ، احداثيات نقاط البداية والنهاية ، عدد العقد ، الاتجاهات من أجل مقارنتها وبيان امكانية دمجها. اقترحت هذه المنهجية لدراسة حالة معينة : استخدام البيانات الجغرافية من خريطة الشارع المفتوح (OSM) وبرنامج Google Earth (GE) كأمثلة على مصادر البيانات المجانية. اظهرت النتائج امكانية تكامل بيانات شبكات الطرق من المصادر المجانية للمعلومات المكانية وتوظيفها لخدمة وتطوير منظومة البنى الأساسية للمعلومات المكانية (SDIs).

1. INTRODUCTION

During the past decade the development of the World Wide Web (www) has led to a massive increase of geospatial information by non-specialist users. This kind of information was termed by **Goodchild, 2007** as volunteered geographic information (VGI). Various VGI applications exist that allows users to upload and browse information in various media (text, pictures, videos, documents, etc.), where such information becomes 'spatial' through links to a spatial reference**, DeLongueville, et al., 2010**. Nowadays, there is a wide variety of geospatial data sources available on the Internet such as the Google Map service, the OpenStreetMap (OSM) project, the Yahoo imagery, the Google Earth (GE) software and many others. Inconsistent data quality is expected in a VGI datasets which makes data integration more challenging task.

Data integration is generally defined as the process of combing different datasets to improve one of them or to enhance the visual presentation of overlaid spatial datasets, **Koukoletsos, et al., 2012**. Integrated geospatial datasets may have different quality issues, often due to the way that each was compiled. Therefore, the data quality issue becomes important when integrating spatial data. A number of authors have considered the effects of geospatial data quality on data integration. For example, **Neis,** and **Zielstra, 2014** studied the current developments in VGI research, focusing on the different methods that were applied to analyze the VGI data quality. They concluded that the development of new methods that compute a trust factor, contributor reputation or individual contributor data quality is required. **Koukoletsos, et al., 2012** proposed an automated feature-based matching method specifically designed for VGI. It is applied to the OSM dataset using the official data from Ordnance Survey as the reference dataset. The results are then used to evaluate data completeness of OSM in several case studies in the UK. In the same view, **Pourabdollah, et al., 2013** presented a method for conflating road attributes, namely the name and reference code, of OSM with the open data provided by Ordnance Survey. It was shown that the best correspondence between attributes exists in the very dense areas, followed by the very low density areas, and lastly in the middle to large sized cities.

The main goal of the current study was to create a program which can assess and report on linear geospatial data quality. This program requires the inclusion of a range of linear data quality measures (e.g. lengths, centeroids, start and end points, number of nodes and directions) which will be coded and tested to ensure the creation of a data quality index. This index will be applied to a number of geospatial data sets and an assessment based on it will be made of the possible data integration of such data sets. The types of data set to be used include OpenStreetMap (OSM) and Google Earth (GE). The result will be a working system or potentially a report on linear data quality, which can be used to integrate data sets.

2. BACKGROUND

2.1 Spatial Data Infrastructure: Development and Concepts

In general, a spatial data infrastructure (SDI) can be defined as the implementation of a framework of geographic data, technology, human resources, software, and polices that are linked in order to process, disseminate and share geospatial data in effective way. Over the last few years, various authors have been defined the term “SDI” differently. For instance, according to **Williamson, 2004** the SDI means a special case of Information Infrastructure, specifically geared towards geographic information. In a follow-up study, **Hjelmager, et al., 2008** wrote that the SDI is an evolving concept about facilitating and coordinating the exchange and sharing of spatial data and services between stakeholders from different levels in the spatial data community. In a study conducted by **Mohammadi, et al., 2010**, it was shown that the SDI aims to better address the technical and nontechnical issues and facilitate data integration. The SDIs aim to provide a holistic platform for users to interact with spatial data through technical and nontechnical tools.

In their analysis of SDI, **Rajabifard,** and **Williamson, 2006** identified the main components of SDI as a data, policy, network, standards and people, as shown in **Fig. 1**. In general, a data component represents various layers of geospatial data. The layers may include satellite images, geographic names, coordinates, administrative boundaries and elevation for a country or jurisdiction. These geospatial data sources are managed by different communities and their geospatial data quality are different as well. For instance, maps for necessary purposes such as land use maps, topographic mapping for the military applications and cadastral maps have been produced by formal or governmental institutions. In many countries, much of these data are protected by the data license due to their high quality. At the same time, the maps which have been performed by these institutions are invested and considered as a source of economy. This can be easily recognised through pricing policies and legal copyright, **Perkins,** and **Dodge, 2008.** Whereas, the recent development in communication technologies allowed for establishing open source collaborative map projects which usually free datasets, **Geller, 2007**. The differences between various sources of free datasets can cause some integration problems for SDI applications. Hence, this investigation has attempted to resolve data integration problem of linear features from OSM and GE sources to serve SDI management.

2.2 OpenStreetMap Project

The OpenStreetMap (OSM) project (www.openstreetmap.org) was started by Steve Coast in England in 2004, and its aim is to create a free world map. In many ways the OSM is similar to Wikipedia, the free encyclopaedia. Anyone can contribute (edit, add or delete data) to Wikipedia, just as anyone can contribute to OSM. The central purpose of OSM is to collect geodata, and make that data available to anybody in its raw form. In addition to that, the project also offers a number of different maps on the web, which are created from this raw data. By looking at OSM one can be able to select different map layer such as Standard, Cycle Map, Transport Map, MapQueqst Open and Humanitarian, **Ramm,** and **Chilton, 2011**.

Objects drawn on an OSM map are called map features. A map features have a geographic coordinates and descriptions. OSM project includes several groups of features such as roads and railways; forests, lakes and rivers; coastline and islands; buildings and land use areas. The most important group of objects in OSM are, of course, streets. They are tagged using the highway key, with a value describing the type of street or way. The road classification in OSM service was based on the road types in the world, as illustrated in **Table 1, Ramm,** and **Chilton, 2011.** In this analysis, the road network of all OSM ways was used for integrating with the correspondence ways of GE datasets.

2.3 Google Earth Software

Google Earth (GE) is the most popular 3-dimension virtual globe system that offers free access to high resolution imagery for most of the global. It was launched in 2005 by Google. The GE can also provide digital elevation data, which is collected by NASA’s Shuttle Radar Topography Mission. In recent years, the using of GE has become more popular by general public and scientific communities. This is due to the fact that the GE is user friendliness, free of charge, wide availability, minimal system and computing power requirements. In addition, GE do not need for extensive training compared with popular geographic information systems (GIS) software, **Benker, et al., 2011**. One of the most gorgeous features of GE is the use of one coordinate system for the whole of the world, that is, WGS84 (World Geodetic System which was found in 1984). This means that every position on the Earth can be stored by the value of geographic coordinates (latitude / longitude). Therefore, GE has been used, for example, to collect ground control points (GCPs), for orthorectification of satellite imagery, to estimate urban vegetation cover, to visualize the output of scientific experiments, and as reference data to evaluate land cover datasets, **Hernández, et al., 2013**.

Although these are positive aspects, however, the resolution of GE images is vary for different areas of the globe. Besides, the 3D viewer modelling is limited for some countries and cities, for instance there is no 3D data for Iraq, while other places such as UK have a detailed 3D data for all cities and countryside. Another limitation of GE is that the inconsistency of positional accuracy of GE data. Therefore, a series of accuracy assessments of GE's imagery have been undertaken by different researchers. For example, **Benker, et al., 2011** assessed the accuracy of GE imagery in the Big Bend region of Texas, USA based on high precision field measurements (<1 m). They found that GE's imagery has a horizontal mean error of 6.95 m and Root Mean Square Error (RMSE) of 2.64 m. In an investigation into accuracy assessment of GE data, **Mohammed, et al., 2013** estimated the GE horizontal and vertical accuracy in Khartoum state. The finding showed that the (RMSE) for horizontal coordinates was 1.59 m and for height measurement was 1.7 m. In another major analysis, **Pulighe, et al., 2015** examined the horizontal accuracy of very-high resolution GE images for the Rome / Italy. The results revealed that very-high resolution GE images have an overall positional accuracy to 1 meter.

To the best of our knowledge, there is a lack of evaluating the integration of GE data with other open source data; therefore this research attempts to contribute to the presenting of a methodology for integrating linear features, such as roads, form GE and OpenStreetMap (OSM) data sources.

3. STUDY SITE AND DATASET

The study area of this case study is situated in the city of Al-Jadriya, placed in Baghdad the capital of Iraq **(Fig. 2)**. It covers approximately 18 km^2. The study area covers few centres of population (including the city centre), and some scattered green areas around them. The reason for selecting this area was to test the possibility of integrating open source geospatial data outside the cities where there are high levels of participation such as UK. The selected site is located within the longitude ranging from 44.368694° E to 44.436500° E and the latitude ranging from 33.262502° N to 33.311578° N. From the map of the study area can see the details of the roads that were under analysis in the research (see **Fig. 2 and Table 1**).

In this study, the data of OpenStreetMap (OSM) service were downloaded from Geofabrik (www.geofabric.de) in 2015. This server has data extracts from the OSM project which are normally updated every day. Data on the download server is organised by region. The OSM data directory contains files that have a whole continent's data in them, and for some continents there are subdirectories in which you find individual files for various countries. Some countries again have their own subdirectories with data for administrative subdivisions. For instance, the OSM data for Iraq are available as a one directory with different layers for the whole country. The Geofabrik service can provide OSM data as shapfiles which can be processed directly by almost GIS software.

In order to enable a comparative analysis, the Google Earth (GE) roads dataset was used as other open source data. The GE makes it easy to search for the coordinates of any location and zooming in on them. There are many ways to obtain coordinates: move the mouse cursor to where you want, and note the coordinates displayed in the lower left-hand corner of the GE display window. Other useful way to obtain coordinates from GE is creating a place mark (probably in the temporary folder). Right-click or control-click on the place mark; one of the options is "Copy". Copy the place mark and paste it back to a text editor. The result is a KML (Keyhole Markup Language) description of the place mark, and since KML is XML, it's easily manipulable. The latest method has been adopted to acquire GE data for the study area of this research.

Fig. 3 illustrates a sample of the road networks in OSM and GE. Red lines represent GE data and blue lines represent OSM data. It can be seen from the figure below that the roads from two data sources are almost similar. However, in some places there are some roads recorded in GE data whereas they are not recorded in OSM data. Therefore, adopting a convenient data integration process can improve and enrich OSM data based on GE data or vice versa.

4. PROGRAM DESIGN

A geospatial data from different sources contain various features. Data integration of multisource geospatial data, take advantage of the strengths of a single data source for improvement of visual interpretation and quantitative analysis. In order to achieve the main goal of this project (data integration), the properties of tested data must be checked carefully. It is indispensable because the information of different free sources is often not complaint to any standard and each organisation is producing the datasets at various level of richness. There are different methods of properties quality assessment and several pieces of research have focused on evaluating them, **Tveite, 1999; Devillers, et al., 2006; Haklay, 2010**. In this investigate the assessment of the quality of data provided by open sources will depend on a comparison of the roads network parameters such as lengths, centeroids, start and end points, number of nodes and directions. The result will be a set of operators which can measure data quality, allow for the preparation of datasets prior to successful integration, and actually undertake the integration of data. This can be used to determine whether the integrated procedure can be achieved or whether the selected data should be changed. This will allow for purposes such as data updates between datasets, contribute to change detection strategies, and assuring map auditing for optimum data quality.

In order to implement the methodology of this research, a program was encoded using Matlab. The Matlab was adopted because it has the ability for producing and analysing scientific graphs, developing effective algorithms, and achieving efficient mathematical computations, in addition to involving tools for programming matrices operations. **Fig. 4** shows a diagram for the workflow of the designed program. In the first step, the program required calling or loading tested data. The tested roads network data extracted using ArcGIS software, and then exported from ArcGIS to save it as .xlsx in order to import into the coded program.

The open source data was designed to include different feature types such as roads, politics boundaries, buildings, lakes, rivers, etc. Therefore, in pre-processing phase the tested data was filtered to obtain only road data for integration test. The purpose of the primary filter is to quickly create a subset of the data and reduce the consuming of the processing time. Afterwards, it is required to enter the number of observations (number of tested roads) into the program. Subsequently, the program will check the integration parameters (conditional expression) such as lengths, centeroids, start and end points, number of nodes and directions to decide the proceeding into integration process or not. If the tested data pass the conditional expression, the program will go directly into successful integration, otherwise the program will return back into the previous step to select alternative dataset.

5. EXPERIMENT AND RESULTS ANALYSIS

5.1 The Roads Length

The intensity of mapping activities of open source data for any city can be reflected by road length characterization. Therefore, it is necessary to consider roads length as one parameter of roads network data integration processing. The assumption is that the length of compared features is approximately similar to the length of its matching feature(s). In this research, the roads length was imported from OSM and GE databases using ArcGIS 9.3. The roads lengths of 212 pairs of homologous roads were considered. It can be seen from **Fig. 5** that more than 70% of the differences in the length are within 5m. This is quite a good number for the matching length of roads from OSM and GE datasets.

5.2 The Centroid of the Roads

In mathematics and physics, the centroid or geometric centre of a two-dimensional region is the arithmetic mean (average) position of all the points in the shape. The definition extends to any object in n-dimensional space: its centroid is the mean position of all the points in all of the coordinate directions. Informally, it is the point at which a cardboard cut-out of the region could be perfectly balanced on the tip of a pencil, assuming uniform density and a uniform gravitational field. The centroid of any polyline (X_c , Y_c) can be defined as follows:

$$X_c = \frac{1}{\int_l dL} \int_L x\, dL \quad (1)$$

$$Y_c = \frac{1}{\int_l dL} \int_L y\, dL \quad (2)$$

Where dL is the differential element of length, and it can be expressed as $d_L = \sqrt{dx^2 + dy^2}$. In this project, the centroid of the selected pairs of roads was calculated and the results are presented in **Fig. 6**. As can be seen from the bar chart below, there are no instances of the shift between the centroid of OSM and GM datasets lay between zero and less 3m, and nearly 70% of the shift lay between 3m and 9m and the rest has spread over a wide range up to 17m. Hence, this is an indication of an efficient judgment of roads integration from OSM and GE data sources.

5.3 Start and End Points

Regarding the accuracy of start and end points of roads, efforts to date have focused primarily on road intersections only, **Zielstra,** and **Zipf, 2010; Haklay, 2010; Girres,** and **Touya, 2010**, and have not evaluated the accuracy of other point features such as start and end points of roads. In order to expand the current state-of-knowledge, this part of study has addressed point features representing the start and end points of the roads of OSM and GE datasets. **Fig. 7 and 8** illustrate the summary statistics for the differences between the coordinates of start and end points of the compared roads. It is clear from the data in charts that the distribution of both groups is similar. In addition, over sixty percent of those lay between 0 and 6 m which indicated that this comparison can be a good metric for integrating OSM and GE datasets.

5.4 Number of Nodes and Vertices

Shape points are the ends and bends that define the feature’s outline. At the beginning and end of every line feature is a node, whereas at each bend (change of direction) is a vertex. A shape of any feature is recorded by using the coordinates of its shape points. In this study, the difference in node number of all OSM and GE object lines were determined. As shown in **Fig. 9**, almost two-third of the compared roads (68%) has the same number of nodes and vertices, while only 28% has one node different, which is quite acceptable for integrating linear features.

5.5 The Angle of Road Direction

A recent study by **Zhao, et al., 2015** suggested that the using of road direction can be a good measure for linear data integration. In this research, the measuring of the angle of direction was prepared according to the procedure used by **Zhao, et al., 2015**. Thus the road angles were measured based on a planer coordinates system joining the start and end points of the line and the vertical axis. **Fig. 10** shows the summary statistics for the preliminary analysis of the differences in the direction of roads between tested datasets. It is apparent from this data that nearly eighty percent of the differences in direction were within 5 degrees. According to these observations, we can infer that it is possible to choose 5 degrees as threshold values for integrating network roads from OSM and GE datasets.

6. DISCUSSION AND CONCLUSION

The free and online mapping sources have experienced increase in geospatial data contributions in recent years. This is due to the fact that profiting from the wide availability of online geo-data on the Internet. The amounts of free geographical information are increasing every day. For instance, a recent statistics showed that the geospatial information database project provides

1400 web map servers that provide over 200,000 map data layers. Profiting from the free accessibility of UGC data (licence, cost, sharing) and opening up a new paradigm of geo-data-people's SDI. Thousands of users are able to access to current volunteered geographic information (VGI) data without any charge whereas; in European countries formal geographical data is expensive and out of the reach of individuals, **Perkins,** and **Dodge, 2008**. Potential benefits in integration e.g. contributing to more informed decision and making videos of location. Nowadays, in any part of the world, the VGI sources let visitors to search a world map and download different portions remotely, **Haklay,** and **Weber, 2008**. The VGI content can also potentially fill in the gaps in places where formal data is scare, unreliable and unavailable. Hence, huge benefits can be gained by integrating open source datasets especially in developing SDI applications.

Conventionally, the geospatial data for an SDI have provided by official (governmental) mapping agencies. However, users usually pay to obtain official mapping data; therefore VGI becomes another way to contribute for spatial information management and SDI development. Today, the VGI is one of the embedding areas within SDI especially for applications need to update SDI quickly such as emergency services. For instance, the United Nations based on VGI from OSM as the most important source to collect geospatial data and create route maps when an earth quick hit Haiti in January 2010. Therefore, it is necessary to encourage volunteers to enlarge the free and open source maps and overcome the potential and limitations of their contributions.

The aim of this research was to investigate the integration of open source geospatial datasets and assess its quality. The method compared geometric properties of road networks such as lengths, centeroids, start and end points, number of nodes and directions. The majority of the tests were based on OpenStreetMap (OSM) and Google Earth (GE) datasets. The results showed that more than 70% of the differences in the length are within 5m. Also, there are no instances of the shift between the centroid of OSM and GM datasets lay between zero and less 3m, and nearly 70% of the shift lay between 3m and 9m and the rest has spread over a wide range up to 17m. In addition, over sixty percent of the differences in start and end points lay between 0 and 6m. Moreover, almost two-third of the compared roads (68%) has the same number of nodes and vertices, while only 28% has one node different. The comparison between the directions of the tested datasets reported that almost eighty percent of the differences in direction were within 5 degrees. The findings of this study suggest that in general it is possible to choose the obtained metrics as a threshold values for successful integration of network roads from OSM and GE data sources.

REFERENCES

- Benker, S.C; Langford, R.P.; Pavlis, T.L., 2011, *Positional Accuracy of the Google Earth Terrain Model Derived from Stratigraphic Unconformities in the Big Bend Region, Texas*, USA. Geocarto International, 26(4), PP. 291-301.

- DeLongueville, B., G. Luraschi, P. Smits, S. Peedell, and T. De Groeve., 2010, Citizens as sensors for natural hazards: a VGI integration workflow. Geomatica, 64(1), PP. 41-59.

- Devillers, R. and R. Jeansoulin (eds), 2006, Fundamentals of Spatial Data Quality. Great Britain: Antony Rowe Ltd, Chippenham, Wiltshire.
- Geller, T., 2007, *Imaging the World: The State of Online Mapping*, IEEE Computer Graphics and Applications, March/April: 8-13.

- Girres, J., and G. Touya., 2010, *Quality Assessment of the French OpenStreetMap Dataset*, Transactions in GIS, 14 (4), PP. 435–459.

- Goodchild, M.F., 2007, *Citizens as Sensors: The World of Volunteered Geography*, Geojournal, 69 (4), PP. 211-221.

- Haklay, M., 2010, *How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets*, Environment and Planning B: Planning and Design, 37 (4), PP. 682–703.

- Haklay, M., and Weber, P., 2008, *OpenStreetMap: User-Generated Street Maps*, IEEE Pervasive Computing 7 (4): 12-18.

- Hernández, C., Salinas-Castillo, W., Guevara-cortina, F., and X. Martínez-Becerra, 2013, *Horizontal Positional Accuracy of Google Earth's Imagery over Rural Areas: A study Case in Tamaulipas, Mexico*, BCG - Boletim de Ciências Geodésicas.

- Hjelmager, J. et al., 2008, *An Initial Formal Model for Spatial Data Infrastructure*, International Journal of Geographical Information Science, 22(11), PP.1295-1309.

- Koukoletsos, T; Haklay, M; Ellul, C, 2012, *Assessing Data Completeness of VGI through an Automated Matching Procedure for Linear Data*, Transactions in GIS, 16 (4), PP. 478 - 498.

- Mohammed, N., Ghazi, A., and Mustafa, H., 2013, *Positional Accuracy Testing of Google Earth*, International Journal of Multidisciplinary Sciences and Engineering, 4 (6), PP. 6-9.

- Mohammadi, H., Abbas, R. and Ian P. Williamson, 2010, *Development of an Interoperable Tool to Facilitate Spatial Data Integration in The Context of SDI*, International Journal of Geographical Information Science, 24 (4), PP. 487–505.

- Neis, P. & Zielstra, D., 2014, *Recent Developments and Future Trends in Volunteered Geographic Information Research: The Case of OpenStreetMap*. Future Internet, 6, PP. 76- 106.

- Perkins, C., M. Dodge, 2008, *The Potential of User-Generated Cartography: A case Study of The OpenStreetMap Project and Mapchester Mapping Party*, North West Geography, 8, PP. 19-32.

- Pourabdollah, A.; Morley, J.; Feldman, S.; Jackson, M., 2013, *Towards An Authoritative OpenStreetMap: Conflating OSM and OS OpenData National Maps' Road Network*, ISPRS Int. J. Geo-Inf., 2, PP. 704–728.

- Pulighe Giuseppe, Valerio Baiocchi & Flavio Lupia, 2015, *Horizontal Accuracy Assessment of Very High-Resolution Google Earth Images in The City of Rome, Italy*, International Journal of Digital Earth

- Rajabifard, A. and I. Williamson, 2006, *The Integration of Built and Natural Environmental Datasets Within National SDI Initiatives*, Seventeenth United Nations Regional Cartographic Conference for Asia and the Pacific, New York**:** 7.
- Ramm, F., Topf, J. and Chilton, S., 2011, *OpenStreetMap - Using and Enhancing the Free Map of The World*, Cambridge, England: UIT Cambridge Ltd.
- Tveite, H. and S. Langass, 1999, *An Accuracy Assessment Method for Geographical Line Data Sets based on Buffering*, International Journal of Geographical Information Science, Vol. 13 (1): PP. 27-47.
- Williamson, I., 2004, *Building SDIs—The Challenges Ahead*, In Proceedings of the 7th International Conference: Global Spatial Data Infrastructure, 2–6 February, Bangalore, India.

- Zielstra, D., and A. Zipf., 2010, *A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany*, In 13th AGILE International Conference on Geographic Information Science.

- Zhao, P., Jia, T., Qin, K., Shan, J., and Jiao, C., 2015, *Statistical Analysis on The Evolution of OpenStreetMap Road Networks in Beijing*, Physica A, 420, PP. 59-72.

Table 1. Road classifications and descriptions by OpenStreetMap project, **Ramm,** and **Chilton, 2011**.

Road type	Description
highway, motorway	A restricted access major divided highway, normally with 2 or more running lanes plus emergency hard shoulder. Equivalent to the Freeway, Autobahn, etc..
highway, trunk	The most important roads in a country's system that aren't motorways. (Need not necessarily be a divided highway.)
highway, primary	The next most important roads in a country's system. (Often link larger towns.)
highway, secondary	The next most important roads in a country's system. (Often link smaller towns and villages.)
highway, tertiary	The next most important roads in a country's system.
highway, residential	Roads which are primarily lined with and serve as an access to housing.
highway, service	For access roads to, or within an industrial estate, camp site, business park, car park etc.

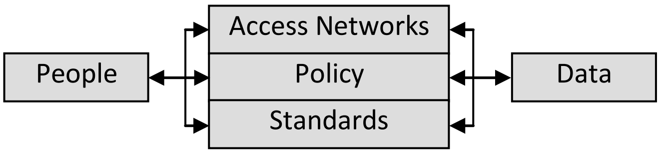

Figure 1. SDI components, **Rajabifard,** and **Williamson, 2006.**

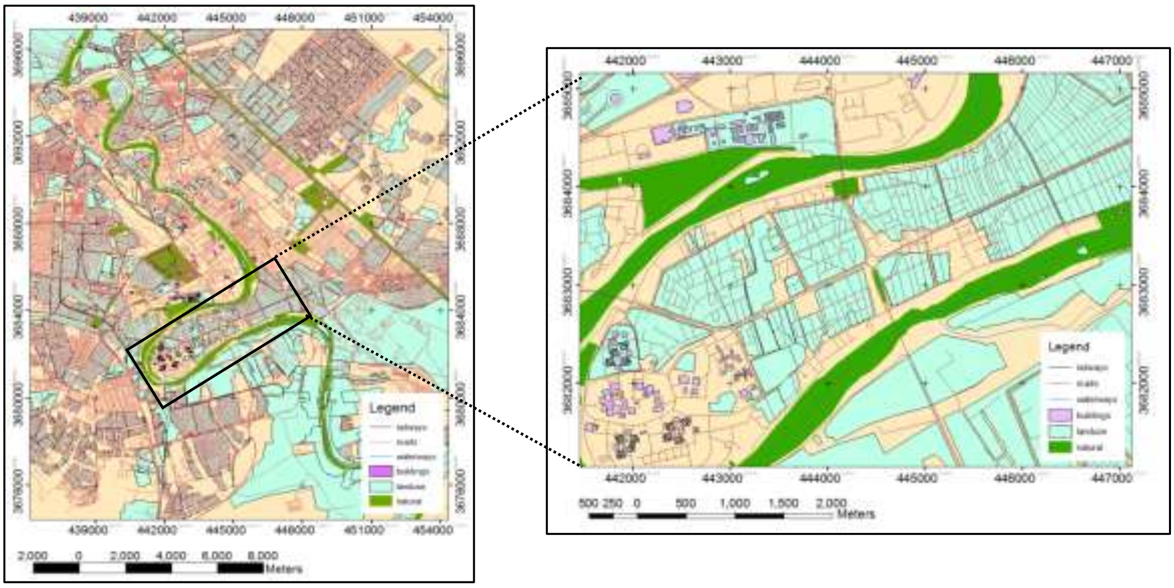

Figure 2. Geographical location of the study area with the studied roads.

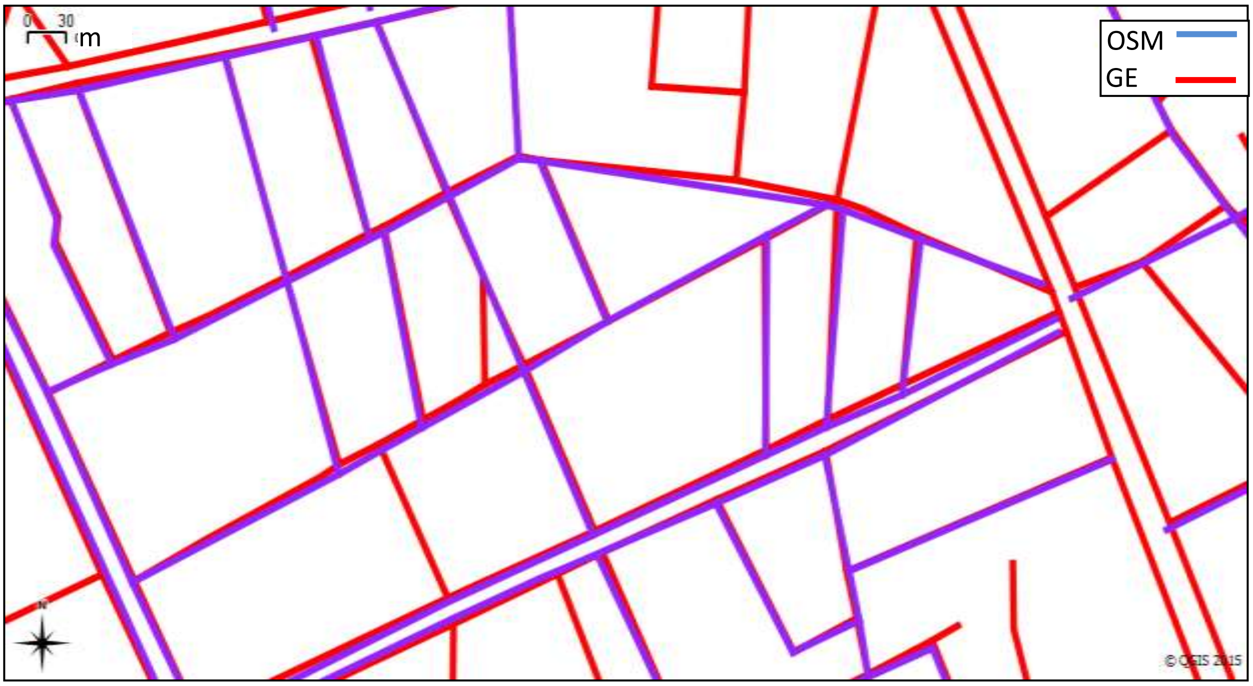

Figure 3. Example of matched and mismatched data.

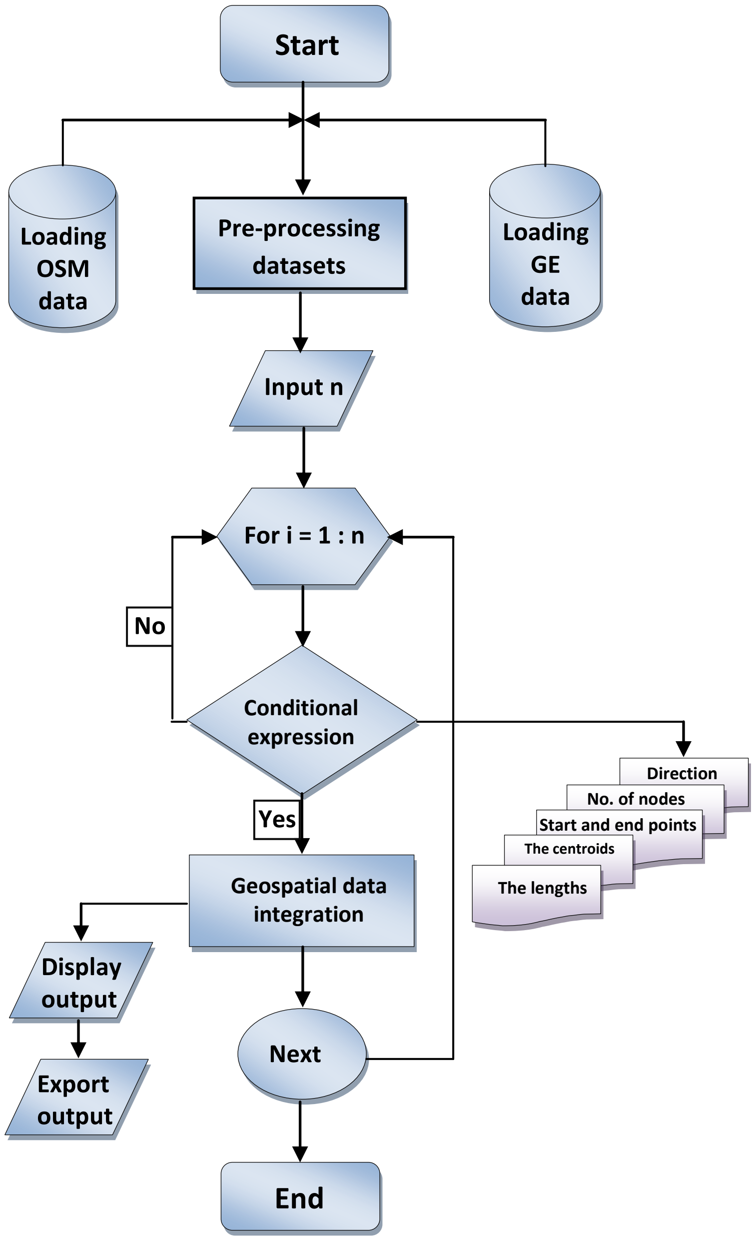

Figure 4. The workflow of the designed program

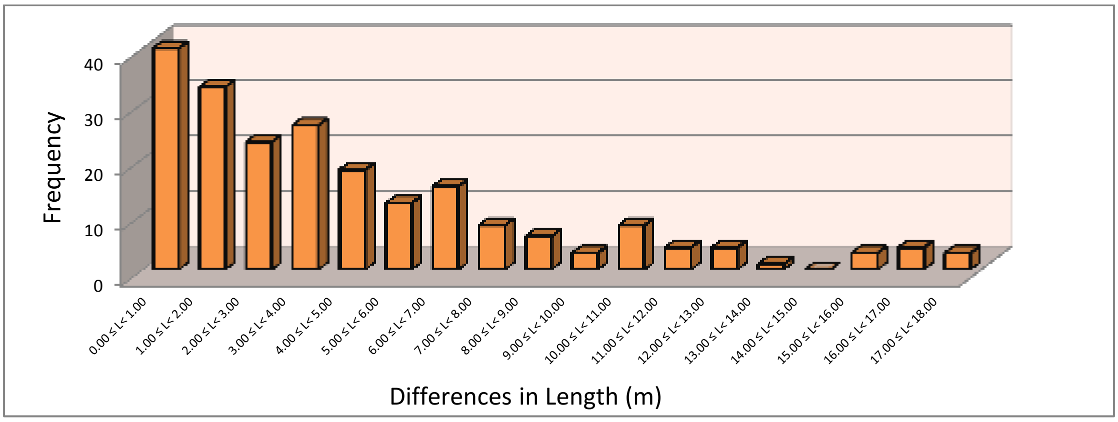

Figure 5. The diffrences in roads length.

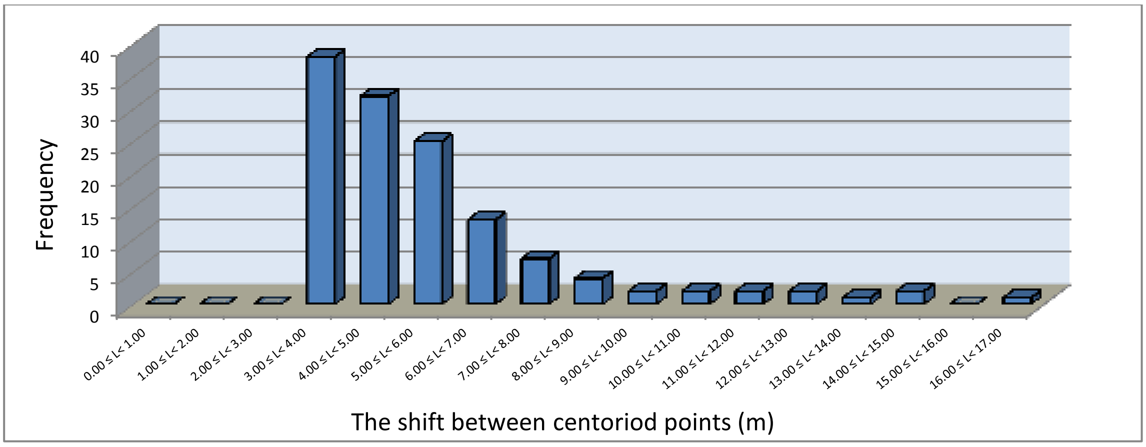

Figure 6. The differences in centoroid.

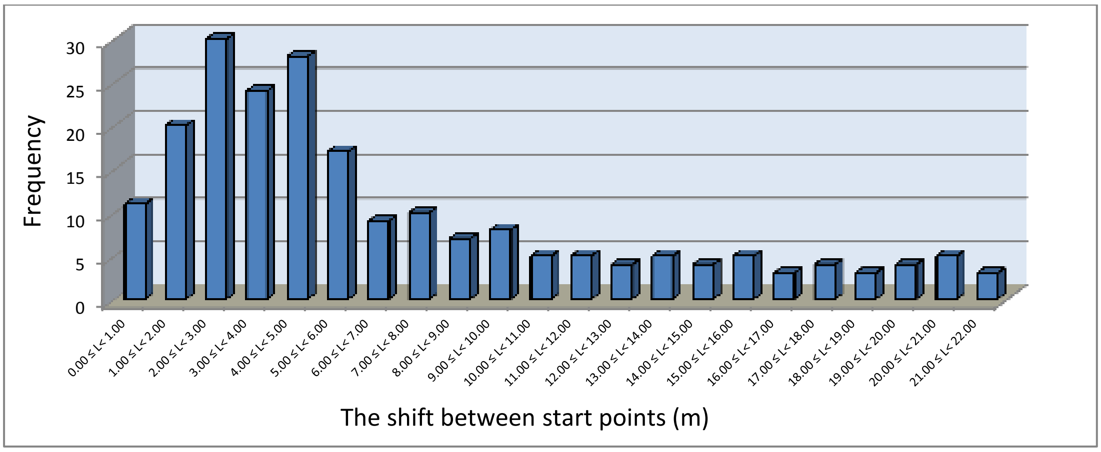

Figure 7. The differences in start points.

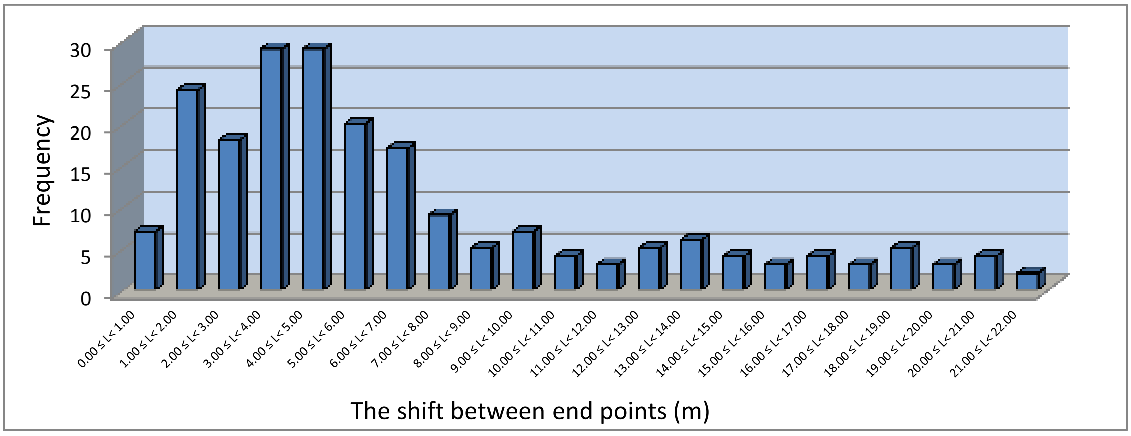

Figure 8. The differences in end points.

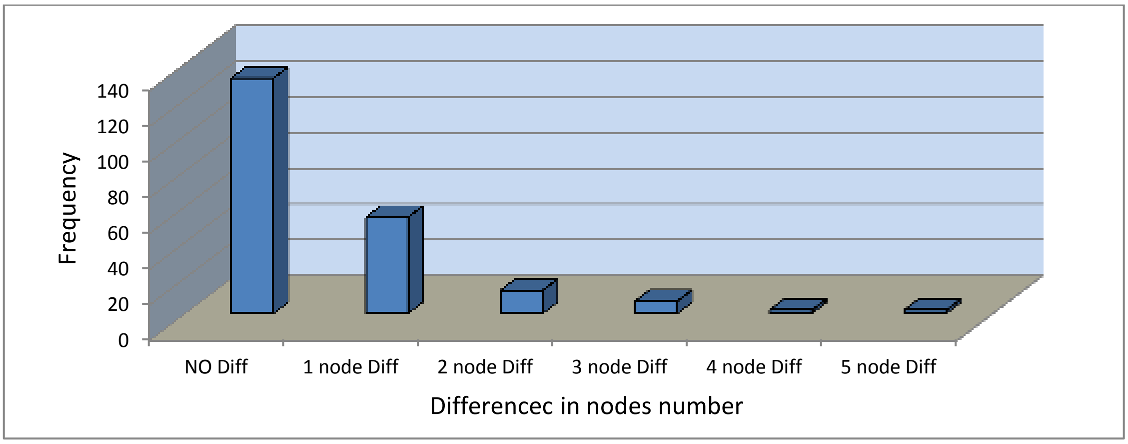

Figure 9. The differences between node numbers.

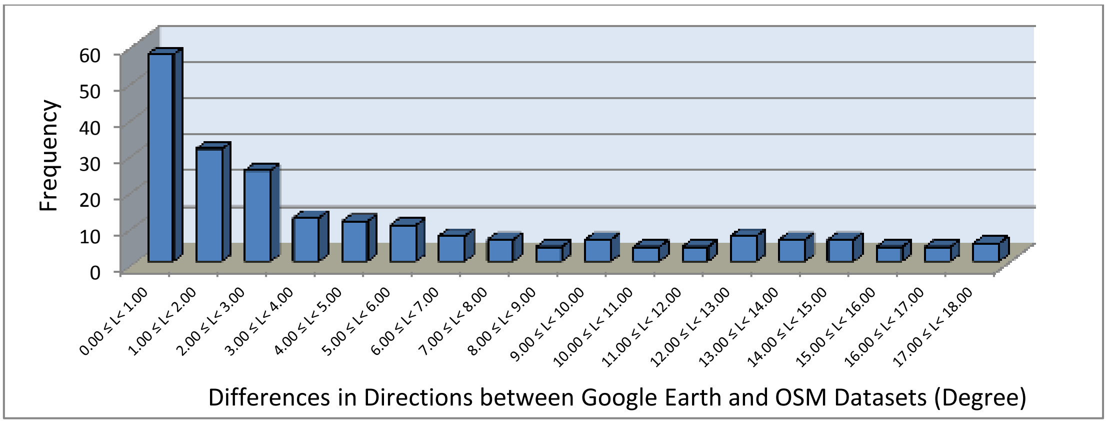

Figure 10. The differences in direction.