SINGLE LAYER MAP VIEWER APPROACH FOR MULTI-SCALE DATA EXTRACTION

WAN MUHD HAIRI BIN WAN AB KARIM

UNIVERSITI TEKNOLOGI MALAYSIA

SINGLE LAYER MAP VIEWER APPROACH FOR MULTI-SCALE DATA EXTRACTION.

WAN MUHD HAIRI BIN WAN AB KARIM

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Geoinformatics)

Faculty of Built Environment and Survey
Universiti Teknologi Malaysia

MAY 2022

DEDICATION

This thesis is dedicated to family, friends, and supervisors who taught me that the best kind of knowledge is that which is learned for its own sake. It is also dedicated to my wife and daughters (Aisyah and Sarah), who struggled to complete this thesis. Last but not least, to I Net Spatial Sdn Bhd's team members that worked together to complete the projects during the pandemic and my study leave.

ACKNOWLEDGEMENT

In preparing this thesis, I was in contact with many people, researchers, academicians and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor, Professor Sr. Dr. Alias Bin Abdul Rahman, for his encouragement, guidance, criticism and friendship. I am also very thankful to my co-supervisor Dr Nor Suhaibah Binti Azri and my ex-supervisor Dr. Pawel Boguslawski for their guidance, advice and motivation. Without their continued support and interest, this thesis would not have been the same as presented here.

I would like to acknowledge MyBrain15 from the Ministry of Higher Education for funding this study at the beginning (3 years) and also to the faculty staff members who were very friendly and helped me in the research.

My fellow lab-mate and the officemate should also be recognised for their support. My sincere appreciation also extends to all my colleagues and others who have provided assistance at various occasions and in the consultation projects. Their views and tips were useful indeed. Unfortunately, it is not possible to list all of them in this limited space. I am grateful to all my family members who patiently are waiting for my PhD completion.

Lastly, my thanks to the Department of Survey and Mapping Malaysia, Map2u Sdn. Bhd. and a few agencies providing datasets for this research.

ABSTRACT

In any desktop and online mapping platform, almost all types of spatial datasets are prepared in multi-scale levels to support map ratio in each zoom class. Current practices in multi-scale data management are either by file-based, database, layer tiles or by using generalisation algorithms to support each transition zoom level of online maps. Multi-scale datasets, either 2-Dimension (2D) and 3-Dimension (3D) vectors or raster, are prepared to serve as a set of data in several accuracy levels or level of details (LoDs) for sharing purposes. However, existing solutions have several drawbacks, either at the storage cost, updating workload or visualisation (graphic and memory), and are time-consuming. For example, vector datasets lead to redundancy in geometry, attribute, topology and semantics for each LoDs and are unable to be viewed in a single viewer. On the other hand, the raster dataset has multiple resolutions from various sources and times, leading to extra storage and slower rendering performance. Thus, this study aimed to develop new approach in properly managing existing multi-scale spatial data such as vector 2D, 3D CityGML and rasters toward a single layer map viewer. The study introduced the Scale Unique Identifier (ID) for vectors to connect all respective LoDs in the attribute database, enabling cross-LoD information query. The High Definition (HD) Map Extractor tool was developed for the raster, and the Enhanced Terrain Profile (ETP) was upgraded based on works in QGIS software to support cross-resolution queries respectively for polygons and lines. The performance of each mentioned dataset was tested, especially in machine resources utilisation on memory, graphic and processor for updating workload, time taken for cross-scale query, and cost-benefits compared to the existing solutions. The experiment performed in this study improved up to 75 percent of time taken for information retrieval, cost-benefits and maintenance efficiency compared to the existing solutions. The findings would benefit data owners and providers in sharing their spatial datasets while minimising the drawbacks. The study has proven that proper construct, control and management of multi-scale spatial datasets would undoubtedly encourage and expedite data sharing among respective data owners, agencies, stakeholders and public users. The study could be extended by improving data sharing standards, implementing scale unique ID in spatial databases and single viewer for 3D city models.

ABSTRAK

Dalam mana-mana *desktop* dan platform pemetaan dalam talian, hampir semua jenis set data spatial disediakan dalam beberapa tahap berskala bagi menyokong nisbah peta dalam setiap kelas zum. Amalan semasa dalam pengurusan data berskala adalah sama ada berdasarkan fail, pangkalan data, lapisan berjubin atau dengan menggunakan algoritma generalisasi untuk menyokong setiap tahap peralihan zum pemetaan dalam talian. Set data berskala, sama ada vektor 2-Dimensi (2D) dan 3-Dimensi (3D) atau *raster*, disediakan bagi berfungsi sebagai satu set data dengan beberapa tahap ketepatan atau tahap perincian (LoD) untuk tujuan perkongsian. Walau bagaimanapun, penyelesaian sedia ada mempunyai beberapa kelemahan, sama ada pada kos penyimpanan, mengemas kini beban kerja atau visualisasi (grafik dan memori), dan memakan masa. Sebagai contoh, set data vektor membawa kepada lebihan dalam geometri, atribut, topologi dan semantik bagi setiap LoD dan tidak boleh dipaparkan dalam satu paparan. Sebaliknya, set data *raster* mempunyai pelbagai resolusi daripada pelbagai sumber dan masa, yang membawa kepada tambahan penyimpanan dan prestasi pemaparan yang lebih perlahan. Oleh itu, kajian ini bertujuan untuk membangunkan satu pendekatan baru dalam menguruskan dengan betul data spatial berskala sedia ada seperti vektor 2D, 3D CityGML dan *raster* ke arah paparan peta lapisan tunggal. Kajian ini memperkenalkan skala unik *Identifier* (ID) bagi vektor untuk menghubungkan semua LoD dalam pangkalan data-atribut, membolehkan pengekstrakan maklumat antara LoD. Alat *High Definition Map Extractor* (*HD Map Extractor*) telah dibangunkan untuk *raster*, dan *Enhanced Terrain Profile* (RTP) telah dikemas kini berdasarkan kerja dalam perisian *QGIS* bagi menyokong pengekstrakan antara resolusi berdasarkan poligon dan garisan. Prestasi setiap set data yang disebutkan telah diuji, terutamanya dalam penggunaan mesin pada memori, grafik dan pemproses bagi beban kerja pengemaskinian, masa yang diperlukan bagi pengekstrakan maklumat antara skala dan manfaat kos yang dibandingkan dengan penyelesaian sedia ada. Eksperimen yang dilakukan dalam kajian ini telah meningkatkan sehingga 75 peratus masa yang diambil untuk mendapatkan maklumat, manfaat kos dan kecekapan penyelenggaraan berbanding dengan penyelesaian sedia ada. Hasil kajian ini akan memberi manfaat kepada pemilik dan penyedia data dalam berkongsi set data spatial mereka disamping meminimumkan kelemahan. Kajian telah membuktikan bahawa cara pengumpulan, pengendalian dan pengurusan set data spatial berskala yang tepat akan mendorong dan mempercepatkan perkongsian data di kalangan pemilik data, agensi, pihak berkepentingan dan pengguna awam. Kajian ini akan diteruskan dengan menambah baik piawaian perkongsian data, pelaksanaan skala unik ID dalam pangkalan data spatial dan pemapar tunggal untuk 3D model bandar.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

GIS	-	Geographic Information System
IT	-	Information technology
2D	-	2-Dimension
3D	-	3-Dimension
CAD	-	Computer Aided Design
nD	-	N^{th} Dimension
SDI	-	Spatial Data Infrastructure
LoD	-	Level of Detail
LADM	-	Land Administration Domain Model
JUPEM	-	Jabatan Ukur dan Pemetaan Malaysia
RAM	-	Random Access Memory
BLG	-	Binary Line Generalisation
ADE	-	Application Domain Extension
RGB	-	Red Green Blue
LiDAR	-	Light Detection and Ranging
DTM	-	Digital Terrain Model
QGIS	-	Quantum GIS
ID	-	Identifier
KM	-	Kilometre
UAV	-	Unmanned Aerial Vehicle
RTK	-	Real Time Kinematic
BAC	-	Building Area Coverage
MBH	-	Mean Building Height
RPV	-	Remote Piloted Vehicle
JPEG	-	Joint Photographic Experts Group
POI	-	Point of Interest
imp	-	importance
BLGtree	-	Binary Line Generalisation tree
SSC	-	Space Scale Cube
OGC	-	Open Geospatial Consortium

IFSAR	-	Interferometric Synthetic Aperture Radar
DEM	-	Digital Elevation Model
SRTM	-	Shuttle Radar Topography Mission
DBMS	-	Database Management System
HTML	-	HyperText Markup Language
TIFF	-	Tag Image File Format
IMG	-	Image format
LZW	-	Lempel-Ziv-Welch
I/O	-	Input Output
BLOB	-	Binary Large Object
API	-	Application Programming Interface
LBS	-	Location-based Services
AR	-	Augmented Reality
MS1759	-	Malaysia Standard for Features and Attribute Codes 1759
MLS	-	Mobile Laser Scanning
UPI ID	-	Unique Parcel Identifier
XML	-	Extensible Markup Language
ISO	-	International Organisation for Standardization
OODB	-	Object-Oriented Database
HD	-	High Definition

LIST OF SYMBOLS

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Nowadays, the Geographic Information System (GIS) is widely used to capture, store, manipulate, analyse, disseminate, retrieve spatial data as referenced geographical information (Dueker, 1989) and present (modelling/visualising) as an effective decision-making tool. Decision- makers had shifted and relied on GIS since early 1990, especially to solve GIS- related complex spatial problems via integrating knowledge on database, analytical model, tabular reporting, visualisation capability and others. Demands of various spatial data information are essential especially for decision- makers in specific and related domains.

In a successful implementation of the GIS system or application, a combination of several basic components should coexist. There are five well-known basic components - data, people (expertise or/and user), procedure, hardware and software. Absence of any of the mentioned components during implementation results in unreliable spatial information or, in the worst case, failure in the GIS system/project as a decision-making tool. One of the mentioned GIS components which is considered as the first priority in setting up any GIS analysis, units, systems or applications – the data. GIS data (formally called spatial data) is the main factor in differentiating either a particular system based on standard Information Technology (IT) or the GIS system. A systematically managed GIS dataset is very essential and undoubtedly will increase performance of four other components such as people (less workload), procedure (easy method), software (functionality, analysis and viewer), hardware (selection and cost).

As for GIS data, there are two categories of spatial data types; vector and raster datasets (Maffini, 1987). Later, they are subdivided into dimensionality levels such as well-known terms used across professional domains. For example, spatial dataset covers vector and raster for two-dimensional (2D) dataset while 3D buildings and terrains are normally called 3D models. For vector data type, it can be in multiple dimensionality - nD (where n is from 0-3 at the moment for spatial components), comprising of geometry, topology, semantic integration and attribute components, while the raster data type only has single values such as elevation, density or other parameters. However, any spatial dimension dataset/model may have different time frames and representation details which are covered in this thesis.

Recently, the evolution and popularity of the GIS spatial data are quite advanced, especially in the aspects of modelling platforms (software) and hardware development (survey instruments, computers and related Internet technology). Traditionally, spatial data was presented in two-dimensions (2D) such as printed maps, whereas these days three-dimension (3D) GIS has become a trend for spatial data modelling (Al-Hanbali et al., 2006). A good set of GIS dataset/model, while utilising the integration of procedure, hardware and software components, undoubtedly produces crucial information describing real -world phenomenon and man-made features for related organisations/governances as a geo-decision tool.

As for 3D modelling, software components, especially from commercial and open-source platforms offer built-in functionalities. For instance, Computer Aided Design (CAD), Revit, Blender, Sketchup, FreeCAD, Paraview, Transmagic and others provide 3D modelling and visualisation solutions which introduce several 3D data formats. Although 3D modelling is considered as a norm, some functionalities (analysis) and queries are still in the early stage and very limited as compared to the overall capability in the 2D spatial datasets environment (Karim et al., 2017). Thus, many basic and complex functions or analysis available in the 2D GIS can hardly be implemented in the 3D GIS environment, for example, spatio-temporal map generalisation (Mioc et al., 2013). Geometry, topology, semantic integration and feature attributes are the basic keys of spatial data requirements in serving GIS data-ready especially for n-dimensional (n>1) data model presentation (Karim et al., 2016).

The previous paragraph is all about spatial dimension (2D and 3D) for both the vector (geometry, topology, semantic and attribute) modelling and the raster data type which are considered as mainstream for spatial modelling. However, the GIS trend is shifting towards new dimensions which are classified as non-spatial dimensions such as scale (e.g. level of details model/information), time (e.g. time frame/temporal such as historical data) and other parameters (e.g. temperature, sound etc.) depending on the needs of the GIS-related domains and applications. The research society is moving forward to integrating a highly formal definition of geo-data (Oosterom and Stoter, 2010), especially since the last decade. Thus, the scale and the temporal dimension have become trends for research and implementation in the 2D and 3D modelling environment. For example, Google Earth services provide temporal imageries data to suit some applications on land use - urban changes, better resolutions which consists a series of orthophoto images of current and previous year. These images are spatially georeferenced to support non-spatial dimensions (e.g. temporal and scale/resolutions) especially toward 'Big Data' preparedness and efficiency (fast, better performance, minimum machine resource consumption and costs) retrieve information (data management) in the future.

In the GIS perspective, a multidimensional (nD) model may consist of one, two, three or more non-spatial dimensions supporting spatial objects (Karim et al., 2016; Gold, 2005). Many approaches, methods, and algorithms (data model or data structure) can be used in defining higher dimensional or hybrid of spatial and non-spatial models. The 3D model possibly may also consist of the 2D geometry of spatial dimensions accompanied by non-spatial dimensions such as scale or time as the third dimension (Oosterom, 2005). This idea is also supported by Gold (2005) who suggests that a 3D model may consist of a 2D geometrical primitive of spatial dimension and unconnected data of objects or parameters. The unconnected geometrical data also can be understood as a viewpoint reflection (observation) of spatial objects (Zhou & Jones, 2001) in different perspectives such as a series of details (multi-scale) or time / temporal dimension. Thus, the 3D multi-scales and 3D-temporal dimension may consist of a 2D spatial object linked to a series of different details or temporal (time) maps which contain mostly identical 2D objects. Therefore, the 3D-scale is the dimension of a series of different representation details (1D non-spatial) of the same 2D mapping objects while the 4D-scale is for the 3D object with multi-representation details.

However, using these 3D term will cause confusion among GIS users (who mostly recognise 3D as spatial X,Y,Z), especially for laymen, clients, professional and system developers who have limited background on the mapping (GIS) concept. Thus, from this section onwards, the terms 2D-scale and 3D-scale which will be used often, refer to the 2D spatial dataset with additional scale dimension, while the 3D-scale term will refer to the 3D spatial model with scale dimension.

Literature on this non-spatial scale dimension shows that since early 2000, researchers have attempted to add another dimension such as scale and time into the spatial GIS modelling (Worboys, 1994; Raper, 2000; Peuquet, 2001) but had less significant results due to unsupported available data structure and data model. Peuquet (2001) and Worboys (1994) focused on the temporal dimension, while Oosterom (2005) and (Li, 1994) focused on the scale dimension. The implementation of the scale dimension faces many problems mostly due to the limitation of available data structures especially for three and higher dimensions. Only a few models extended from the Multi-Scale Line tree (Jones & Abraham, 1986), Arc-tree (Günther, 1988), Binary Line Generalisation (BLG) (Oosterom, 1990) and others are potentially able to integrate geometric and scale aspects in one representation.

Scaling dimension in GIS has gained some popularity in recent years due to the demand from users or applications to extract and share their 2D and 3D data models across mapping levels (macro and micro) which could provide spatial information to various related domains for decision making process. The GIS research community is focusing on integrating a highly formal definition of geo-data (Oosterom & Stoter, 2010) and thus is focusing on designing the most efficient framework for the implementation of the scale dimension. It includes efficient zooming, query, storage (avoids redundancy and inconsistency as much as possible), schema, format, standard levels of accuracy and progressive transfer between the client and the server sites (if possible, for huge online data retrieval). Researchers also started to design and propose a variety of efficient frameworks, implementation and compression techniques, e.g. CitySAC by Siew and Kumar (2019)) and sharing procedures such as schema/standard for the 3D-scale into the targeted application. CityGML is an example of the international standard sharing schema (with LoDs) and the Spatial Data Infrastructure (SDI) for each country's operational framework (local schema).

Since different applications and users need specific details of data representation, Sester (2007) suggested that different representations or different Level of Details (LoDs) of the same reality have to be made available (with or without explicit relationships between corresponding features at the adjacent LoDs). Thus, there is a need to combine all LoDs into a single container called scale dimension. Combining all LoDs into a systematic structure while minimizing redundancy, supporting information retrieval from each level, reducing cost for updating/maintenance and enabling data-sharing across domains will be the major concern and focus for scale dimension.

Traditionally, 2D multi-scale datasets are stored independently without any connection with other LoDs. This method has several drawbacks mainly on the redundancy of storage (geometry, attribute and topology) and updating time/effort for multiple detail layers on the same object. It slows the performance in utilising spatial datasets as the time taken will get longer (more iteration of queries) either in the data base or the GIS software, especially during visualisation and querying information. Currently, scale dimension for the vector can be categorised into two main approaches: multi-scale and vario-scale, while the raster data type has more than five popular categories which will be discussed in the next chapter.

As for visualisation, the 2D viewer in the desktop-based software is still unable to support multiple scale datasets (retrieval of other information) within a single view, while updating works for a particular object is done manually for each LoD, either by file-based or spatial database. The 2D-scale dataset does not have any relational relationship (semantic) with other LoDs of the same viewed object for visualisation purposes (except for the vario-scale implementation, which will be discussed in Chapter 2). Thus, users could not retrieve other detailed information for geometry, attribute and topology. On the other hand, the 2D online map (using any web map service) has the capability to visualise the 2D-scale dataset, but only by using the on-off layers (shows and hides all scale layers at a certain map scale ratio) technique within certain zoom levels (in/out range).

1.2 Problem Statements

Applied definitions or terms of scale dimension might vary according to user's background, field of interests, needs, software use and system. The Scale-Space Theory in Computer Vision (Lindeberg, 1994), the Scales and Cross-scale Dynamics Dimension in social–ecological systems (Vervoort et al., 2012) and the Multi-scale Fractal Dimension in pattern recognition and image analysis are some examples of the scale theory and the need for a scale dimension in serving their working model or profession. In geoscience, a new multi-scale method incorporating fine-scale information to a coarse-scale equation (Aarnes et al., 2007) and multi-scale 3D visualisation (Jones et al., 2009) have gained popularity since the past decade.

In modelling real-world objects, GIS also considers the scale dimension as a very essential modelling aspect, especially in presenting the details of the spatial objects. Different detailed spatial objects can either be 2D or 3D geometrical space, highly needed by various professions and applications as a mapping-decision support tool. In GIS theoretical perspective, scale can be considered as a non-spatial dimension of an object/a group of objects presented in different viewpoints (Karim et al., 2016). The viewpoint highly depends on the users' professional background on how they see and interprete the object details with respect to their needs. In general, the three main classifications of detail elements in mapping are geometry, semantic/attribute and topology. For example, different LoDs (viewpoints) in a 2D vector can either be with or without the same details of geometry, or/and attribute, or/and topology or a mixed combination of those three elements with respect to each domain's needs.

These scenarios normally occur when an object/model is shared across different domains and professions with multiple viewpoint details. Sometimes they may happen within the same unit or department of an organisation depending on the designated output information details, different formats, systems/software, and developed applications. For example, a state or a national mapping agency with 3D spatial data that can be modelled into 3 types of models suited for usage, user and developed applications are CityModel (CityGML), Land Administration Domain Model (LADM) and the 3D reality mesh model. All these models are represented in their respective levels of details (LoDs). The definition and specification of each LoD1, LoD2 and LoD3 representations in CityGML are different in LADM LoD schema. They are also different in 3D mesh LoD1 (low resolution), LoD2 (medium resolution), LoD3 and so forth.

The scenario described above is also supported by Talhofer et al. (2018) where the vector thematic characteristics (e.g. attribute) are unlimited in numbers and it is possible for each spatial object to own all the necessary thematic data while the vector topological characteristics enable objects to connect with each other via simple to complex topological networks in supporting different analysis including the shortest and the fastest route. These topological networks have various levels, e.g. from the simplest topology to complete topology as mentioned by Talhofer et al. (2018) for each respective level, e.g. level 1 – Spaghetti, level 2 – Chain Note, level 3 – Planar Graph and level 4 – Full Topology (with respect to further study). Specific attributes and topology will also be big constraints for data- sharing between users. Thus, they need to be standardised.

The current implementation frameworks for scale integration (newly discussed term as dimension) within a spatial object, mostly at local and national levels either use generalisation techniques or store the individual level of detailed data into separated databases. An example is the United Kingdom Ordnance Survey and the Malaysia National Mapping Agency (JUPEM) use a series of scale ratios to serve several applications. Through this technique, 2D topographic data at predefined scales such as 1:5,000, 25,000, 1:50,000 and 1:100,000 are stored in separated databases. These data will be called within a single application (e.g. web-mapping portal) which utilises the “on-off layer” technique as a result of users zooming in/out of the map viewer which indirectly consuming higher machine resources such as memory, data streaming (internet download), graphic rendering and low speed.

These recent years have shown an incremental trend on implementing 3D spatial models as compared to past decades. As for the 3D model with multiple scale representations, most countries have not yet engaged with such national LoD specifications and schema. In the early stage of implementation, the Open Geospatial Consortium (OGC) introduced international standards for 3D modelling, e.g. city modelling using the CityGML schema (with four LoDs to represent a 3D building). It is very essential for efforts in standardizing interchange formats, schema, LoD details/requirements for the purpose of effective data-sharing between organisations/domains and a clear direction for GIS software development. This standard also indirectly speeds up research progress and implementation upon 3D models for interested countries as pioneers in advanced mapping (better understanding and accurate decision reflecting real-world phenomenon). However, several drawbacks need to be understood and minimised, especially for creating and maintaining the LoD dataset.

There are three major drawbacks for the existing methods serving multi-scale data (implementation). A predefined scale undoubtedly produces storage redundancy and "heavy" consumption on computer resources such as the Random-Access Memory (RAM). RAM needs to read and store several pre-defined data at different scales and on/off layer based on the zoom level within the visualisation application. Thirdly, this approach causes a lot of upcoming problems such as difficulty in updating (needs more work to update certain objects for every pre-defined scale layer), introducing more errors during the updating process and not being able to preserve geometry consistency/originality after a few times of updating the data (versioning). Thus, it indirectly will be slowing down the data-sharing initiative by the agencies at state, national or international levels.

These different detail level (LoD) specifications and requirements actually can be varied based on the level of the GIS users, data owners and applications. It could be implemented in several data formats, schemas and viewed in various viewers or applications. However, some problems will arise especially in preserving the LoD for attribute and topology, interchange format and supported application for sharing data across domains and users. Local and international standards such as CityGML only cover limited LoDs as compared to vast LoD details as required by each user. Thus, several users stored most detailed data and engaged in generalisation techniques that suit their operations. Generalisation is a technique to aggregate, simplify and transmigrate from detailed data to a lower detailed level. The generalisation technique is normally used to reduce storage and updating works for respective LoD data (geometry, attribute and topology). The generalisation algorithms vary for 2D and 3D spatial data based on priority performance, either to preserve geometry, topology or semantics/attributes information. Many researchers embarked on this subject since early 1990 for 2D (e.g. Binary Line Generalisation (BLG) and Reactive Tree by Oosterom, 1990), and matured in recent years as described by Meijers (2011), particularly in the vario-scale model research. The vario-scale model uses the 2D smooth zoom mapping technique (using generalisation) in 2D scale dimension. As for the 3D generalisation, it has become a research trend since early 2007, where several promising techniques were introduced such as the Three-step Strategy by Baig et al. (2013).

Generalisation reduces storage cost for 2D and 3D vector data, but consumes the processing core (central processing unit, CPU) of the machine. However, it does not overcome the existing drawbacks especially for updating (data maintenance) for each LoD. Attribute and semantic information are also not affected by the generalisation algorithm (except when using the Application Domain Extension, ADE), which means there is no way to generalise the raw tabular information. Updating the attribute of a certain LoD in generalisation is also not working as perfectly as multi-scale data could have provided. Thus, there should be a general solution to utilise the existing LoD data and systematically link with the viewer and information such as the integration with a database.

On the other hand, the raster spatial data type also has the same problems in managing the multi-scale raster. Multi-scale for raster is represented by different resolutions upon an object for each layer. Unlike the vector spatial data type, the raster data type does not have any topology, semantic and multiple attribute information. Each spatial raster (pixel) only has a single parameter, for example elevation and density, except for imagery (combination of Red, Green and Blue colours, RGB).

However, literature (Karim et al., 2019) shows that there is less research work and implementation on this subject except to speed up the visualisation rendering by using tiling, pyramid, compression and other methods. These techniques introduce more raster layers (tiles of resolution levels) and mostly cause higher storage consumption, data streaming, machine memory and graphic.

As a nutshell of problem statement, this research intends to discuss the available solutions on scale dimension modelling for the vector (2D and 3D) and the raster spatial data type with the intention to overcome the mentioned limitations respectively. The research proposes new approaches for respective multi-scale datasets (vector and raster) while utilising the existing structure and LoD of the spatial data. The proposed solutions will be described in details in Chapter 3.

1.3 Research Questions

As discussed in the problem statement in Section 1.2, the main research questions are listed below:

i. What is the current status, limitations, requirements and major drawbacks of the existing multi-scale spatial datasets management?
ii. How to design better solutions for each vector (2D and 3D model) and raster multi-scale datasets while preserving the users' LoD in a single map viewer?
iii. How to test the solution with vector 2D dataset and 3D building models (CityGML)?
iv. How to test the performance of the proposed new approaches in the raster data type?

1.4 Research Aim

This research aims to design new approaches for the efficient management of each multi-scale data type (vector 2D, 3D and raster) to support cross-scale query in a single map viewer and single data layer (less details). The research investigates several existing multi-scale datasets approaches and highlights several drawbacks, then demonstrates the proposed solutions for each spatial data type and dimension. Workability and performance based on results and cost-effective (workload, manpower, and time taken for maintenance) are evaluated as well for the consumption of hardware resources (RAM, GPU) during the query.

1.5 Research Objectives

Based on the research questions described in Section 1.3, the objectives are:

1. To review the current implementation and available solutions for supporting multi-scale dataset management.

2. To design and develop new approaches vector (2D & 3D) and raster technique for utilising existing multi-scale GIS datasets.

3. To demonstrate the workability of the multi-scale GIS data management via a single layer map viewer concept.

1.6 Research Scope

The study utilises the existing multi-scale datasets for both the vector (2D and 3D) and the raster. The vector data comes from the 2D land use dataset and the 3D CityGML (LoD0, LoD1, LoD2 and LoD3) from The Department of Survey and Mapping Malaysia (JUPEM). On the other hand, the raster data comes from orthophoto imagery and Digital Terrain Models (DTM) – sourced from Light Detection and Ranging (LiDAR) from JUPEM. It covers a part of Selangor, approximately 6.25 km2 in area. However, limitation on the availability of building interior data as for LoD4 (indoor measurement of building) and others, indirectly will constrain the scope of the research. Thus, this research only addresses the following aspects in spatial data management – processing time, updating workloads, and storage consumption. Other aspects of 2D and 3D data management such as data structure and new concept of LoDs (e.g. LoD2.5, 3.5, or beyond) are out of the scope.

The developed multi-scale information retrieving tool is only for the 2D raster datasets using the Quantum GIS (GQIS viewer), while vector 2D and 3D CityGML datasets are linked with the respective LoDs by using the introduced scale unique ID in the attribute tables and the PostgreSQL database.

1.7 Significance of Study

This research contributes to the scale dimension domain by providing new approaches on how to store and manage a series of existing multi-scale datasets efficiently. The solutions are valid for both vector (2D and 3D) and 2D raster data types. Most of the available solutions are in the vector domain and less in the raster domain, but lead to redundancy in storage, maintenance workload and visualisation performance. However, this study proposes a solution to utilise the current available multi-scale data for both the vector (2D and 3D) and the raster. This contributes to research and industrial (practitioner) domains such as mapping, agricultural, urban planning application, hazard and others which have multiple accuracy and scale ratio (e.g. land use data, satellite data and DTM) for sharing and are supported with a single map viewer (layer). The proposed solution maintains the state-of-art of spatial data as structured in the existing data owner/custodian; without forcing them to engage with national or international standard LoD for data -sharing purposes. For the vector, only minimum chances (one new attribute column) need to be added to engage with the solution.

The study also attempts to prove that there is no single or complete solution supporting the scale dimension but it varies accordingly to the nature of the industry domain, users' needs and spatial data types. It is due to different applications or users that need different LoD, and their definition for the detailed level is not the same for others. Thus, the study tries to tackle this scenario by proposing new approaches in managing the existing and available multi-scale data efficiently (update, utilise, performance). It is an honor if this study later becomes a reference for future research and an implementation guideline in the industry or for GIS practitioners dealing with scale dataset.

1.8 Thesis Structure

This thesis is structured accordingly within 5 chapters. Chapter 1 delivers the background to the study and research problems to be solved in both the vector and the raster spatial scale dimension. Issues and the needs to conduct this study are highlighted in the respective sections of the research questions, objectives, scopes and significance of the study.

Chapter 2 describes in detail the available methods in supporting scale dimension for both the vector and the raster datasets. Multi-scale (pre-defined level) and vario-scale (generalisation) are examples of available scale dimensions in vector form, while the pyramid, compression, resampling, mosaic and tiling/cataloging are examples from raster. A comparison study is conducted to have a general overview on the different benefits offered by each mentioned method, indicating their main focus of the dataset and potential users. Several drawbacks are also highlighted to tackle the remaining problems and the need for another general solution as a designing new approaches on for both the vector (2D/3D) and the raster spatial data types.

In Chapter 3, the general concept on proposed new approaches are presented for both the 2D-3D vector and the raster data types. Characteristics of the proposed solution, the area focusing for improvement compared to the existing solution, and advantages are described in this chapter. The main chapter content is the designs and implements proposed solutions for respective existing multi-scale spatial datasets (vector and raster) systematically to support a single map viewer. It allows readers to comprehend scale dimension and differentiate these solutions from others as described in the previous Chapter 2. For example, in the 3D vector, a conceptual framework is proposed for implementation using the CityGML LoD and database integration which is the new contribution in scale data management. The chapter also describes methods of measuring the performance of the proposed solutions.

Chapter 4 discusses the results and analyses the respective approaches. The implementation and test are based on multi-scale vector (2D data and 3D CityGML model) and raster data types. An example is the implementation of the 3D CityGML model for the city model is based on the proposed best procedure in dealing with multiple LoDs. Some discussions on the results tested in selected areas such as the 3D city model and the performance are presented. Evaluation on raster extraction (analysis) also discusses the two cross-scale information extraction analysis using the line and polygon mask. Extraction (query) and update multi-scale vector data are also discussed using the Postgres database. Lastly, the performance of the proposed solutions as compared to the existing solution are presented, especially on the raster.

Finally, Chapter 5 is the conclusion of the study. Advantages and limitation are discussed in the detail-based implementation of the study area and the hardware used. Potential future research direction and interests are also identified and discussed.

REFERENCES

Aarnes, Jørg, Kippe, Vegard, Lie, Knut–Andreas, and Rustad, Alf Birger. (2007). *Modelling of multiscale structures in flow simulations for petroleum reservoirs geometric modelling, numerical simulation and optimization* (pp. 307-360): Springer.

Anselin, L., and Getis, P. (1993). Spatial statistic analysis and geographic information systems. In *Geographic information systems, spatial modelling*, *and policy evaluation.* (Eds.), Fisher, M. and Nijkamp, P. Springer-Verlag, New York, 1993. Page 35 – 49.

Anselin, L, (1990). What is special about data? Alternative perspectives on spatial data analysis. In *Spatial statistics, past, present and future.* Griffith, (Eds.), Institute of Mathematical Geography, Ann Arbor. 1990, page 63-77.

ArcGIS Enterprise (2019). Online portal. https://gisportal.gov.si/portal/portalhelp/en/server/latest/get-started/windows/what-is-image-service-caching-.htm

Alsubaie, N., and El-Sheim, N. (2016). The feasibility of 3D point cloud generation from smartphones. ISPRS Archives, Volume XLI-B5, 2016 XXIII ISPRS Congress, 12–19 July 2016, Prague, Czech Republic. DOI:10.5194/isprs-archives-XLI-B5-621-2016

Al-Hanbali N., Fadda E., and Rawashdeh, S. (2006). Building 3D GIS modeling applications in Jordan: Methodology and implementation aspects. In: Abdul-Rahman A., Zlatanova S., Coors V. (Eds.). Innovations in 3D geo information systems. Lecture Notes in Geoinformation and Cartography. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-36998-1_37

Baig, S.U., and Rahman, A.A. (2013). A three-step strategy for generalization of 3D building models based on CityGML specifications. *GeoJournal* 78, 1013–1020 (2013). https://doi.org/10.1007/s10708-013-9475-0

Boguslawski, P., 2011. Modelling and analysing 3D building interiors with the dual half-edge data structure, University of Glamorgan, UK, 1-134 pp.

Brédif, M., Vallet, B., and Ferrand, B., 2015. Distributed dimensonality-based rendering of LiDAR point clouds. *The International Archives of the*

Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-3/W3, 2015. ISPRS Geospatial Week 2015, 28 Sep – 03 Oct 2015, La Grande Motte, France , DOI: 10.5194/isprsarchives-XL-3-W3-559-2015

Chou, T., Liu, C., Yeh., M., & Chen, Y. (2008) Using multiscale spatial data in landslide mornitoring and landuse classification interpretation, *Geographic* I*nformation Sciences,* 14:1, 27-35, DOI: 10.1080/10824000809480636

Cao, S., Weng, Q., Du, M., Li, B., Zhong, Z., & Mo, Y. (2020). Multi-scale three-dimensional detection of urban buildings using aerial LiDAR data, GIScience & Remote Sensing, 57:8, 1125-1143, DOI: 10.1080/15481603.2020.1847453

Dabiri, Z. and Blaschke, T. (2019). Scale matters: A survey of the concepts of scale used in spatial disciplines. *European Journal of Remote Sensing* (Taylor & Francis). 2019, VOL. 52, NO. 1, 419–434. DOI: 10.1080/22797254.2019.1626291

Dueker, K. J., Kjcrne, D. (1989). *Multipurpose cadastre: Terms and definitions*. Falls Church. VA:ASPRS and ACSM

El-Mekawy, M. (2010). *Integrating BIM and GIS for 3D city modelling (The case of IFC and CityGML*). Royal Institute of Technology (KTH), Stockholm, Sweden.

ESRI. (2020). ArcGIS Desktop Portal. Online Portal (Access on May 2021). https://desktop.arcgis.com/en/arcmap/10.3/main/manage-data

ESRI. (2020b). ArcGIS Desktop Portal. Online Portal (Access on May 2021). https://desktop.arcgis.com/en/arcmap/10.3/tools/environments

Fan, H., Meng, L., Jahnke, M. (2009). Generalization of 3D buildings modelled by CityGML. In *Advances in Science*. Proceedings of Agile Conference, Hannover, Germany. 2-5 June 2009. pp. 387-405.

Feng, Y., Thiemann, F., and Sester, M., 2019. "Learning Cartographic Building Generalization with Deep Convolutional Neural Networks" ISPRS International Journal of Geo-Information 8, no. 6: 258. https://doi.org/10.3390/ijgi8060258

Gaffuri, J. (2012). *Toward web mapping with vector data.* International Conference on Geographic Information Science, September 2012. DOI: 10.1007/978-3-642-33024-7_7

GeoServer. (2020). GeoServer Online Portal. https://docs.geoserver.org/stable/en/user/tutorials/imagepyramid/imagepyramid.html

Gold, C. M. (2005). Data structures for dynamic and multidimensional GIS. In *Proceedings of the 4th ISPRS Workshop on Dynamic and Multi-dimensional* GIS (pp. 36-41).

Goodchild, M.F., and D.A., Quattrochi. (1997). Scale, multiscaling, remote sensing and GIS. In *Scale in Remote Sensing and GIS.* (Editors). Dale A. Quattrochi, Michael F. Goodchild. CRC Press Inc. p.p. 1-12.

Gröger, G., and Plümer, L. (2012). CityGML – Interoperable semantic 3D city models. *ISPRS Journal of Photogrammetry and Remote Sensing*. 71. 12–33. 10.1016/j.isprsjprs.2012.04.004. Gold, C. M. (2005). *Data structures for dynamic and multidimensional GIS*. Paper presented at the 4th ISPRS Workshop on Dynamic and Multi-dimensional GIS.

Gröger, G., Kolbe, T.H. (2012). *OGC city geography markup language (CityGML) encoding standard.* Open Geospatial Consortium Inc.: Wayland, MA, USA, 2012.

Guo, N., Xiong, W., WU, Q., and Jing, N. (2016). *An efficient tile-pyramids building method for fast visualization of massive geospatial raster datasets*. January 2016 Advances in Electrical and Computer Engineering 16(4):3-8. DOI: 10.4316/AECE.2016.04001

Günther, O. (1988). notes in computer *Efficient structure for geometric data management. #337 in lecture science*. Springer-Verlag, Berlin. Springer Book.

Harrington, P. de B. (2016). *Multivariate curve resolution of wavelet compressed data*. In *Data handling in science and technology*, Volume 30, 2016, Pages 311-332. https://doi.org/10.1016/B978-0-444-63638-6.00009-7

HBS. (2019). https://online.hbs.edu/blog/post/cost-benefit-analysis. Access May 2021.

ICA (International Cartographic Association). (1973). *Multilingual dictionary of technical terms in cartography.* Wiesbaden, Franz Steiner.Jones, C.B., and Abraham, I.M. (1986). *Design Considerations for a scale dependent cartographic database.* 384-398.

Jones, R.R., McCaffrey, K.J.W., Clegg, P., Wilson, R.W., Holliman, N.S., Holdsworth, R.E., Waggott, S., 2009. Integration of regional to outcrop digital

3D data:Visualisation of multi-scale geological models. *Computers & Goesciences*, Volume 35(1), 4-18.

Karim, H., Rahman, A. A., and M. Salleh, M. R., 2019. Multi-scale and scale dimension properties in spatial raster modelling – Concept and current implementation. International Archives of the Photogrammetry. *Remote Sensing and Spatial Information Sciences*, Volume XLII-4/W16, 2019. 6th International Conference on Geomatics and Geospatial Technology (GGT 2019), 1–3 October 2019, Kuala Lumpur, Malaysia.

Karim, H., Rahman, A. A., and Boguslawski, P., Meijers, M., and Oosterom, P.V., 2017. The potential of the 3D dual half-edge (DHE) data structure for integrated 2D-space and scale modelling: A review. *Advances in 3D Geoinformation, Lecture Notes in Geoinformation and Cartograhy*. Springer. DOI 10.1007/978-3-319-25691-7_27

Karim, H., Alias Abdul Rahman, and Pawel Boguslawski. (2016). Generalization technique for 2D+scale DDE data model. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Volume XLII-2/W1, 2016. 3rd International GeoAdvances Workshop, 16–17 October 2016, Istanbul, Turkey.

Karim, H., & Rahman, A.A. (2016). Application perspectiveOF 2D+scale dimension ISPRS - *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 119-126.

Karim, W.M.H.W.A, and Hashim, M.G. (2013). *Development of a new D16 algorithm for single flow direction model*. Book Chapter at Geoinformation for Informed Decisions. Springer 2014. ISBN 3319036440, 9783319036441.

Lam, N.S.-N. (2019). Resolution. *The Geographic Information Science & Technology Body of Knowledge* (2nd Quarter 2019 Edition), John P. Wilson (Ed.). DOI: 10.22224/gistbok/2019.2.11

Lam, N. S. N., Catts, D., McMaster, R. B., Quattrochi, D. A., & Brown, D. (2005). *Scale.* In R. B. McMasters and E. L. Usery (Eds.), *A research agenda for geographic information science*, Chapter 4. Bacon Raton, Florida: CRC Press, pp. 93-128.

Lam, N. S. N., & Quattrochi, D. A. (1992). On the issues of scale, resolution, and fractal analysis in the mapping sciences. *The Professional Geographer,* 44(1), 89-99.

Li, Z,. and Yan, H. (2019). Transformation in scale for continuous zooming. *Manual* , pp 279-324. https://link.springer.com/chapter/10.1007/978-981-32-9915-3_8

Li Z.L., (2008). Multi-scale digital terrain modelling and analysis. In: Zhou Q, Lees B, Tang G. (Eds.), *Advances in digital terrain analysis,* Springer, pp 59–83.

Li, Z.L., 1994. Reality in time-scale system and cartographic representation. *The Cartographic Journal,* 31(1):50-51.

Lindeberg, T. (1994). Scale-space Theory: A basic tool for analysing structures at different scales. J*ournal of Applied Statistic*, Volume 21(2), 225-270.

Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W. (2005). Geographical Information Systems: Principles, Techniques, Management and Applications, 2nd Edition. ISBN: 978-0-471-73545-8. Publisher Wiley.

Maffini, G. (1987). Raster versus vector data encoding and handling: a commentary. Photogramm Eng Remote Sens 53(10):1397–1398

Mao, B., ban, Y., and Harrie, L. (2011). A Multiple Representation Data Structure for Dynamic Visualization of Generalized 3D City Models. ISPRS J. 66(2), 198-208 (2011)

Mioc, D., Anton, F., Gold, C., & Moulin, B. (2013). Spatio-temporal Map Generalizations with the Hierarchical Voronoi Data Structure. ISVD 2013 - 10th International Symposium on Voronoi Diagrams in Science and Engineering. 63-72. 10.1109/ISVD.2013.19.

OGC, (2019). Open Geospatial Consortium. Online portal and Standard. www.ogc.org/

Meijers, M., (2011). Variable-Scale Geo-Information, PhD Thesis. Technische Universiteit Delft, Netherlands.

Oosterom, P.V. and Stoter, J. (2010). 5D Data Modelling: Full Integration of 2D/3D Space, Time and Scale Dimensions, 6th international conference on Geo-graphic information science. Springer- Verlag, Berlin, Heidelberg, pp. pages 310–324.

Oosterom, P.V., Devries, M., and Meijers, M., (2006). Vario-scale Data Server in a Web Service Context. Workshop of the ICA Commission on Map Generalisation and Multiple Representation – June 2006.

Oosterom, P.V., (2005). Variable-scale Topological Data Structures Suitable for Progressive Data Transfer: The GAP-face Tree and GAP-edge Forest. Cartography and Geographic Information Science(32), 331–346.

Oosterom, P.V. and Schenkelaars, V. (1995). The Development of an Interactive Multiscale GIS. Int. J. Geogr. Inf. Syst. 9, 489–507 (1995)

Oosterom, P.V. (1990). Reactive Data Structures for Geographic Information Systems. PhD-thesis Department of Computer Science, Leiden University, December 1990.

Oosterom, P.V., and Meijers, M.. (2014). Vario-scale Data Structures Supporting Smooth Zoom and Progressive Transfer of 2D and 3D data. International Journal of Geographical Information Science. 28. 455-478. 10.1080/13658816.2013.809724.

Pathan, S. K. (2012). Multi Criteria Modelling in GIS. SCRIBD Online Slide. https.//www.scribd.com/presentation/ 20608630/Multi-Criteria-Modelling-in-Gis. Access date. 18 May 2019.

Paul, N., Bradley, P. E., & Breunig, M. (2013). Integrating Space, Time, Version and Scale Using Alexandrov Topologies. International Symposium on Spatial and Temporal Databases SSTD 2013.

Pepe, M., Costantino, D., Alfio, V.S., Angelini, M.G., and Garofalo, A.R. (2020). A CityGML Multiscale Approach for the Conservation and Management of Cultural Heritage: The Case Study of the Old Town of Taranto (Italy). ISPRS Int. J. Geo-Inf. 2020, 9, 449; doi:10.3390/ijgi9070449

Quattrochi, D. A. and Goodchild, M. F. (1997). Scale in Remote Sensing and GIS. Editors: Dale A. Quattrochi and Michael F. Goodchild. Lewis Publishers.

Quiroga C.A., Singh V.P., Iyengar S.S. (1996). Spatial Data Characteristics. In: Singh V.P., Fiorentino M. (eds) Geographical Information Systems in Hydrology. Water Science and Technology Library, vol 26. Springer, Dordrecht. https://doi.org/10.1007/978-94-015-8745-7_4

Peuquet, D.J. (2001). Making Space for Time: Issues in Space-time Data Representation. Geoinformatica, 5: 11-32.

Raj, P. (2017). Raster Data Model. Online - Slide Share. https.//www.slideshare.net/pramodgpramod/raster-data-model-76694132. Access date. 18 May 2019.

Raper, J. (2000). Multidimensional geographic information science. Taylor & Francis, London.

RasterLite (2020). Online RasterLite Manual, Spatial Is Not Special https://www.gaia-gis.it/

Rodrigue, J.P. (2021). GIS Data Model in The Geography of Transport Systems.

Online website (2021). https://transportgeography.org/

Schneider, M. (2009). Spatial Data Types. In: L. Liu and M.T. Özsu (Editors), Encyclopedia of Database Systems. Springer-Verlag.

Sester, M. (2007). 3D Visualization and Generalization. Institude of Cartography and Geoinformatics. Leibniz University of Hannover, Online presentation. http://www.ifp.uni-stuttgart.de/phowo/2007/ presentations/320sester.pdf

Siew, C.B., and Kumar, P. (2019). CitySAC: A Query-Able CityGML Compression System. Smart Cities 2019, 2, 106-117.

Talhofer, V., Hošková-Mayerová, Š., and Hofmann, A. (2018). Quality of Digital Geographic Data and Inforation. Springer Book: Quality of Spatial Data in Command and Control System. 28 Jul 2018 - Technology & Engineering.

Tinghua, A., and Cheng, J. (2005). Key issues of multi-scale representation of spatial data. 30. 377-382.

Vervoort, J. M., Rutting, L., Kok, K., Hermans, F. L.P., Veldkamp, T., Bregt, A. K., and Lammeren, R. V. (2012). Exploring Dimensions, Scales, and Cross-scale Dynamics from the Perspectives of Change Agents in Social–ecological Systems. Ecology and Social, 17(4):24.

View Ranger (2020). (online portal) https://www.viewranger.com/en-gb/world-of-maps/premium-maps?country=gb#step2

Weibel R., Burghardt D. (2008) Generalization, On-the-Fly. In: Shekhar S., Xiong H. (eds) Encyclopedia of GIS. Springer, Boston, MA

Weibel, R., and Dutton, G. (1997). Online publication at School of Geosciences, University of Edinburgh. https://www.geos.ed.ac.uk/~gisteac/gis_book_abridged/files/ch10.pdf

Weinan E. and Jianfeng L. (2011). Multiscale modeling. Scholarpedia, 6(10):11527. doi:10.4249/scholarpedia.11527. Access in November 2020.

Worboys, M.F. (1994). A unified model for spatial and temporal information. The Computer Journal, 37(1): 26-34.

Zalipynis, R A. R., Pozdeev, E., and Bryukhov A. (2018). Array DBMS and Satellite Imagery: Towards Big Raster Data in the Cloud Analysis of Images, Social Networks and Texts, 2018, Volume 10716. Springer Book Chapter. ISBN : 978-3-319-73012-7

Zhou, C., Su, F., Pei, T., Zhang, A., Du, Y., Luo, B., Cao, Z., Wang, J., Yuan, W., Zhu, Y., Song, C., Chen, J., Xu, J., Li, F., Ma, T., Jiang, L., Yan, F., Yi, J., Hu, Y., Liao, Y., Xiao, H. (2020). COVID-19: Challenges to GIS with Big Data, Geography and Sustainability, Volume 1, Issue 1, 2020, Pages 77-87, ISSN 2666-6839, https://doi.org/10.1016/j.geosus.2020.03.005.

Zhou, S., & Jones, C. B. (2001). Multi-Scale Spatial Database and Map Generalisation. ICA Workshop on Map Generalisation 2001, Beijing.