# AN IMPROVEMENT OF SPATIAL DATA QUALITY ASSESSMENT PROCESS ON GEOGRAPHICAL INFORMATION SYSTEM

ARTIESA BINTI ARRIS

A project report submitted in partial fulfilment of the
requirements for the award of degree of
Master of Software Engineering

Advanced Informatics School
Universiti Teknologi Malaysia

JANUARY 2017

*Alhamdulillah.....thanks to Allah who has given inspiration and good health to me throughout for the past of six months.*

***To my beloved husband and children, thank you so much*** *for their patience, courage, love, and endless support during the completion of this thesis. Ibu, love all of you so much!*

***To my supportive mother, Puan Mainar Djlalie,*** *who always motivate me since I was at age 5 and always gave me endless love, trust, constant encouragement over the years, and for her prayers.*

*Dear Dr. Nurulhuda Firdaus Binti Mohd. Azmi who always be my side when I need an idea and opinion for the thesis. Thank you Dr. for being such a wonderful lecturer to me, I am really appreciated and respect!*

*Dear Madam Norizan and Cik Rosminah for letting me the opportunity to be a part of your GIS team for six months. I am really appreciating it! And millions thanks to both of you.*

# ACKNOWLEDGEMENT

In the name of Allah, The Most Gracious, The Most Merciful. All praises to Allah for all the absolute blessing and mercy. Peace upon our missed and beloved prophet, Muhammad and peace be upon his family.

I wish to express my deepest appreciation to all those who giving me their hand to helped me, in one way or another, to complete this project especially to GIS Team at Department of Irrigation and Drainage, Malaysia.

First and foremost I thank God almighty who provided me with strength, motivation, ideas and purpose throughout the project. Special thanks to Academic Mentor Dr. Nurulhuda Firdaus Binti Mohd. Azmi with all her patience, guidance and support during the execution of this project. Through her expert guidance, I was able to overcome all the obstacles that I encountered in these enduring six months of my project. In fact, she always gave me immense hope every time I consulted with her over problems relating to my project.

At last but not least, I would like to take this opportunity to express the thoughtful gratitude from my deep heart to my supportive husband, my lovely children, my mother, and siblings for their love, prayers, and continuous support; both spiritually and materially. You were always there encouraging me up and stood by me through my difficulties time.

# ABSTRACT

The use of Geographical Information Systems has flooded almost every field in the engineering, natural and social sciences, offering accurate, efficient, reproducible methods for collecting, viewing and analyzing spatial data. GIS data have become a progressively essential data type to Malaysian Government Agencies and to business in general. An effective GIS data and information is needed to reduce losses from natural & human-induced disaster as prevention and early warning. Spatial data is one of important GIS element where contain Information about the locations and shapes of geographic features and the relationships between them, usually stored as coordinates and topology. Meanwhile, Spatial Data quality is the degree of data that fulfill the given objective. In other words, completeness of attributes in order to achieve the given task can be termed as Data Quality. Production of data by government agencies as well as by various mapping agencies assesses the data quality standards in order to produce better results. Data created from different channels with different techniques may have a conflict in terms of resolution, orientation and displacement. Spatial Data quality is a pillar in any GIS implementation and application as reliable data are indispensable to allow the user obtaining meaningful results. As a result, study on spatial data quality assessment guideline published by MaCGDI is important to DID and relevant to benchmark. Spatial Data Quality Assessment process flow has been designed in order to improve the existing procedure.

# ABSTRAK

Penggunaan Sistem Maklumat Geografi kini semakin meluas dan telah merangkumi hampir semua bidang antaranya dalam bidang kejuruteraan, sains sosial, menawarkan perkhidmatan yang tepat dan cekap, kaedah pengumpulan data, dan menganalisis data spatial. Data GIS kini, juga merupakan data yang progresif dan penting khususnya untuk sector kerajaan dan sektor swasta di Malaysia. Data GIS amat berguna dan merupakan maklumat yang diperlukan untuk mengurangkan kerugian akibat daripada bencana alam dan juga langkah-langkah pencegahan dan amaran awalan. Data Spatial adalah salah satu elemen terpenting di mana GIS menyimpan maklumat mengenai lokasi, bentuk ciri-ciri geografi dan hubungan antara data tersebut, yakni biasanya disimpan sebagai koordinat dan topologi. Kualiti data spatial adalah tunggak dalam mana-mana pelaksanaan GIS dan aplikasi data yang boleh dipercayai sangat diperlukan untuk membolehkan pengguna mendapatkan hasil yang tepat. Justeru itu, dalam usaha untuk menambahbaik prosidur penilaian kualiti data spatial, kajian mengenai Garis Panduan Penilaian Kualiti Data Geospatial yang diterbitkan oleh MaCGDI telah dilaksanakan. Hasilnya, kajian mengenai Garis Panduan Penilaian Kualiti Data Geospatial telah dilaksanakan dan boleh dipenandaraaskan untuk merekabentuk Proses Penilaian Kualiti Data Spatial kegunaan jabatan secara dalaman. Aliran Proses Penilaian Data Spatial telah direkabentuk bagi menaiktaraf prosedur penilaian data spatial sedia ada.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| DASS | - | Dam Automation Surveillance System |
| DEGIS | - | Darul Ehsan GIS |
| DFG | - | Division of Facilities and GIS, DID Malaysia |
| DID | - | Department of Irrigation and Drainage |
| DIT | - | Division of Information Technology |
| DQ | - | Data Quality |
| DQA | - | Data Quality Assessment |
| GDC | - | DID Geospatial Data Centre |
| GDM2000 | - | Geocentric Datum of Malaysia |
| GDMS | - | Geospatial Data Management System |
| GDQAG | - | Geospatial Data Quality Assessment Guideline |
| GIS | - | Geographical Information System |
| ITRF2000 | - | International Terrestrial Reference Frame |
| JICA | - | Japan International Cooperation Agency |
| MaCGDI | - | Malaysian Centre for Geospatial Data Infrastructure |
| MaCGDI | - | Malaysian Centre for Geospatial Data Infrastructure |
| MPIGIS | - | Master Plan of Integrated GIS Department of Irrigation and Drainage |
| MyGDI | - | Malaysian Geospatial Data Infrastructure |
| NaLlS | - | National Infrastructure for Land Information System |
| PEGIS | - | Penang GIS Database |
| RB-IMS | - | River Basin Information Management System |
| RBMU | - | River Basin Management Unit |
| SDD | - | Software Design Description |
| SOP | - | Standard of Procedure |
| SPDG | - | Sistem Pengurusan Data Geospartial |

# LIST OF APPENDICES

# CHAPTER 1

# PROJECT OVERVIEW

## 1.1 Introduction

Geographic Information System (GIS) is experiencing a rapid growth in Malaysia. GIS has formerly been designed as specialized and complex systems to be used by experts for sophisticated spatial analyses. GIS is a computer-based information system for management, analysis, input and output of geographic data and information. It consists of collection, storage, retrieval, manipulation, analysis, and display of spatially related information (Ondieki & Murimi, 2009) A GIS application consists of a collection of software tools to manage process and display geographical data. The software also facilitate the generation and analysis of maps and serve as an indispensable tool in complex planning and management processes of geospatial analysis. In GIS application, an electronic map is being used widely as the main layer to support decision making in terms of locations geographical features. The quality and accuracy of a map define the value for a GIS Application. Faulty map in GIS application may lead to incorrect or missing information. Thus, it is important to ensure the quality and the accuracy of an electronic map and the geographical information for GIS application.

Spatial data are also known as geospatial data or geographic information is the data (or information) that classifies the geographic location of features and boundaries on earth, such as natural or constructed features, river and more. Spatial data is usually stored as coordinates and topology, can be mapped. Spatial data is often accessed, manipulated or analyzed through GIS. Spatial data consists of lines, points, polygons

and other geographic and geometric data primitives, which can be mapped by location, stored with an object as metadata or used by a communication system to locate end user devices. Geospatial data are an important component for a GIS and GIS software allows user to capture, integrate and analyze different geospatial datasets that have different origins and mostly contains different quality levels. The data qualities directly influence the results of analysis in GIS system and therefore to ensure the data is quality, thus it becomes an essential requirement for data quality assessment to be developed.

GIS data have become a progressively essential data type to Malaysian Government Agencies and other businesses in general. An effective GIS data and information is needed to reduce losses from natural & human-induced disaster as prevention and early warning. Besides that, GIS data will improve human health and well-being, for example GIS for public health. An example is tracking child immunizations, conducting health policy research, and establishing service areas and districts. In addition, from GIS data it will enable to improve governance by providing many benefits to government agencies such as better understanding in daily operation, efficiency and cost savings. GIS data are giving decision makers to make more informed decision for better communication, collaboration inter agencies and coordination. An article from National Geographic News mentioned that, "with GIS, it can helps on crime tracking whereby police track the criminals by its location through cell phone calls, credit card transaction and many more". Previously, detectives used paper maps to track criminals with colour pinpoints and as soon as it is done, it is already out dated. (Anna Brendle, 2002).

## 1.2 Company Background

Department of Irrigation and Drainage (DID) was established in 1972 due to the severe flood occurrences in many parts of West Malaysia in 1971. This department is responsible on flood mitigation, hydrology and coastal engineering. In 2004, DID have moved from the Ministry of Agriculture & Agro-based Industry (MOA) to Ministry of Natural Resource & Environment (NRE). DID has shown to the

development of the country for over 80 years and creating own identity and brand. DID is committed in the sustainable management to ensure smooth development. DID's main duty encompass as stated in Table 1.1:

**Table 1.1:** DID's core business

| Core Responsibility | Brief Description |
|---|---|
| River Basin Management and Coastal Zone | Provide expertise services in river and coastal zone management in an integrated manner including policy making and legislation. |
| Water Resources Management and Hydrology | Manage the national hydrological network that meets world class standard and to provide quality data for operational hydrology, water resources assessment and research. |
| Special Projects | Provide project management services professionally in ensuring the successful implementation of Special Projects designated to the DID. |
| Flood Management | Protect life and property and to reduce flood damages, create a more conducive environment for economic and social development and to give immediate response in facing flood. |
| Eco-friendly Drainage | Manage and regulate the implementation of storm water management plan to reduce the adverse effects of urban storm water on receiving water bodies. |

DID is the only department that is responsible for all water issues in Malaysia. Due to Cabinet Restructuring, the DID, Malaysia which was formerly under the Ministry of Agriculture Malaysia, now operates under the Ministry of Natural Resources and Environment starting 27 Mac 2004. The following Figure 1.1 shows the DID organizational chart effective from 1st August 2009.

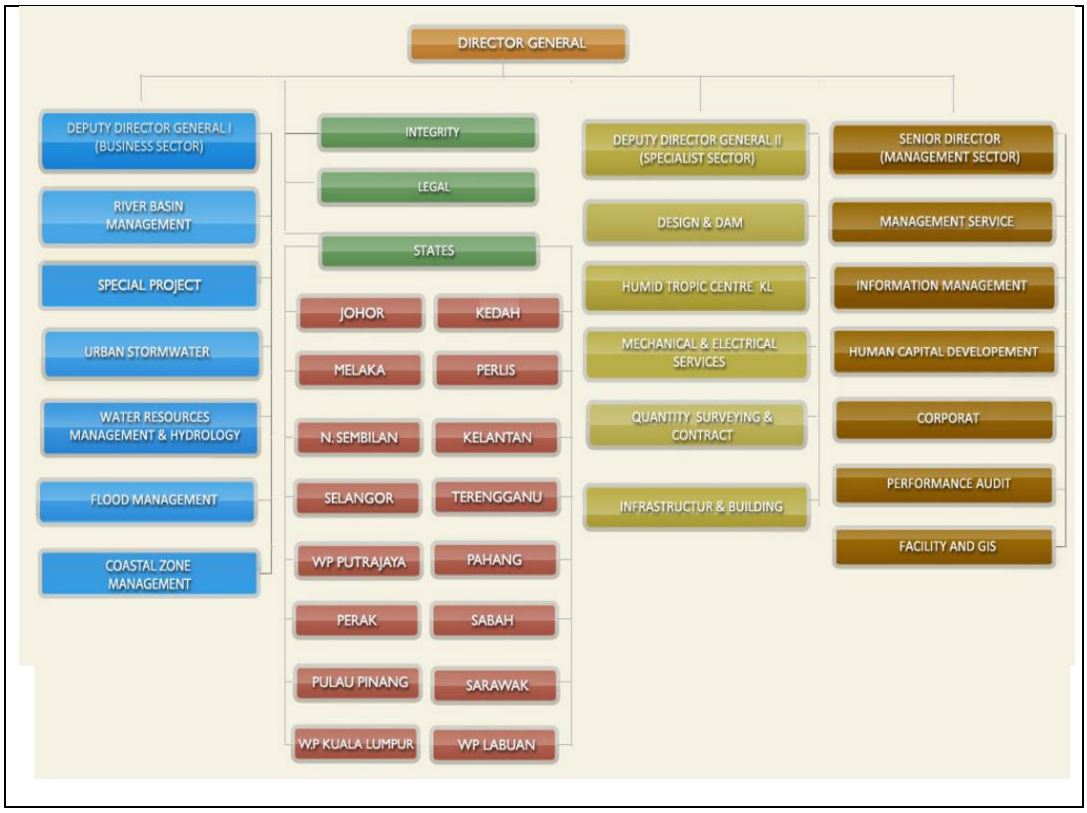


**Figure 1.1** DID's organizational chart

As a public sector organization which related with the management of water resources in Malaysia, DID requires management strategy and integrated GIS model. The GIS model will consists of standard spatial database and the GIS application that will support the core business of DID at federal and state level. Beginning year 2014, DID has established the *Pelan Induk Sistem Maklumat Geografi Bersepadu JPS 2014 - 2018* for a period of 5 years. The plan provide a detail plan on the direction, strategies, sample of procedures, proposals and suggestions of conceptual data model GIS which will serve as a reference for the development of integrated GIS in the DID Malaysia. In addition, this plan will also include the development of *Sistem Pengurusan Data Geospatial* (SPDG) which will be able to facilitate the management and processing of geospatial data for DID.

## 1.3 Project Background

Currently, there are many types of GIS software in the market, used in various applications such as engineering, agriculture, astronomy/planetary, archaeology, architecture, aviation, automobile integration, natural disaster and many more.

In Malaysia many government agencies implement GIS technologies to improve public services such as Flood Warning System, enhance daily operations for example proper traffic management road network, and drive community engagement as for urban planning. In DID, the GIS technologies have been developed since year 1998 as the platform for geospatial analysis which may serve the government to improve public services such as flood forecasting, flood warning, water resources management and hydrology and river basin information and coastal zone management.

SPDG was developed to facilitate the management and to process the geospatial data of DID's. In the preparation of these comprehensive data, the work of the standardization and processing of spatial data should be implemented. This is to ensure that data supply to the related organization either at the federal or state level is reliable, integrity and highly accessible. The processing of geospatial data covers the classification of each layer of data which is based from the DID Geospatial Data Centre (GDC) standard and MS1759:2004 standard. While, managing the geospatial data involve with data structuring according to DID requirements, harmonization of spatial data standards, determining the coordinates of the location and establishment of geospatial database. The development of SPDG consists of 7 modules and this module will be developed by Teknologi KBSE Sdn Bhd. Figure 1.2 shows the list of modules that will be developed in SPDG and Table 1.2 explained briefly about each of the module.

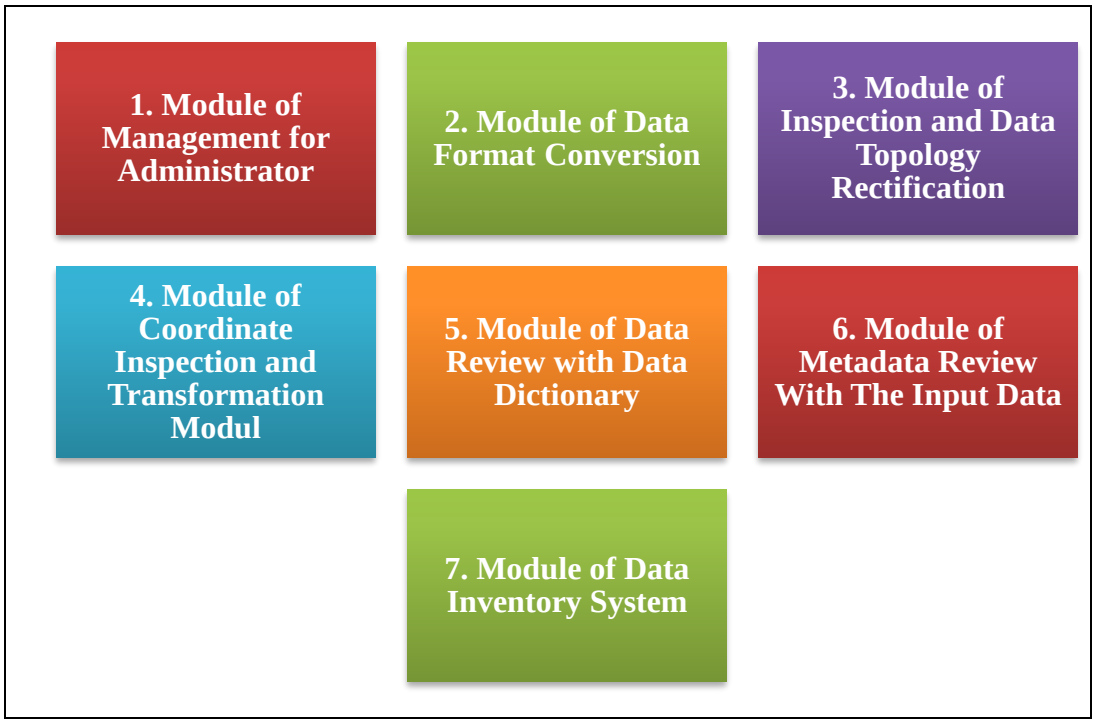


**Figure 1.2** List of modules in SPDG

**Table 1.2:** Brief explanation on SPDG module

| No | Module | Brief Descriptions |
|---|---|---|
| 1 | Management for Administrator | Used to manage user management, display list of user and the audit trail (User daily activities). |
| 2 | Data Format Conversion | Helps users to facilitate the conversion of data format and able to convert various format of data into one standard format. |
| 3 | Inspection and Data Topology Rectification | Able to inspect topological error and repair geometry which involved line, polygon and point. |
| 4 | Coordinate Inspection and Transformation | Check and convert the coordinate system data to a consistent format. |
| 5 | Data Inventory System | Facilitate the searching of spatial data in the database. |
| 6 | Data Review with Data Dictionary | Filtering and reviewing all the data processed in the SPDG and issued a report after the review process completed. |

| No | Module | Brief Descriptions |
|---|---|---|
| 7 | Metadata Review With The Input Data | To check existing metadata information or to create new metadata if not available and also able to edit, create and print metadata for the spatial data. Metadata standard are based on Malaysian Metadata Standard Profile |

## 1.4 Problem Statement

SPDG was initiates from National Flood Forecast and Warning Information Management System (FFWIMS). In SPDG, it is important to ensure that geospatial data used is accurate and follows the format and standards used to create models for flood forecasting. Thus, it is crucial to conduct geospatial data quality assessment in order to make sure that it follows the standard defined by the respective organization. Another important feature for SPDG is the accurateness of the geospatial data used. Accurate geospatial data is important in order to generate result on geospatial data analysis and modelling. From the result it will be used to assists on decision making by the stakeholder. To illustrate in detail about SPDG as part of FFWIMS, GIS component it can be refer to Figure 1.3.

Currently there is no systematic Standard of Procedure (SOP) for spatial data quality assessment in DID. Refer Figure 1.4 for existing SOP. Current practice on geospatial data quality assessment in DID is based on manually identified based on individual skill and experience in GIS technology. Division of Facilities and GIS DFG), was responsible to lead and provide services in GIS to support management of water resources, flood, river basins and coastal zones in DID Malaysia. DFG was establishing to play roles as:

i. To establish, manage and maintain GIS database
ii. To develop and manage appropriate GIS applications
iii. To enhance DID staff's competencies in GIS

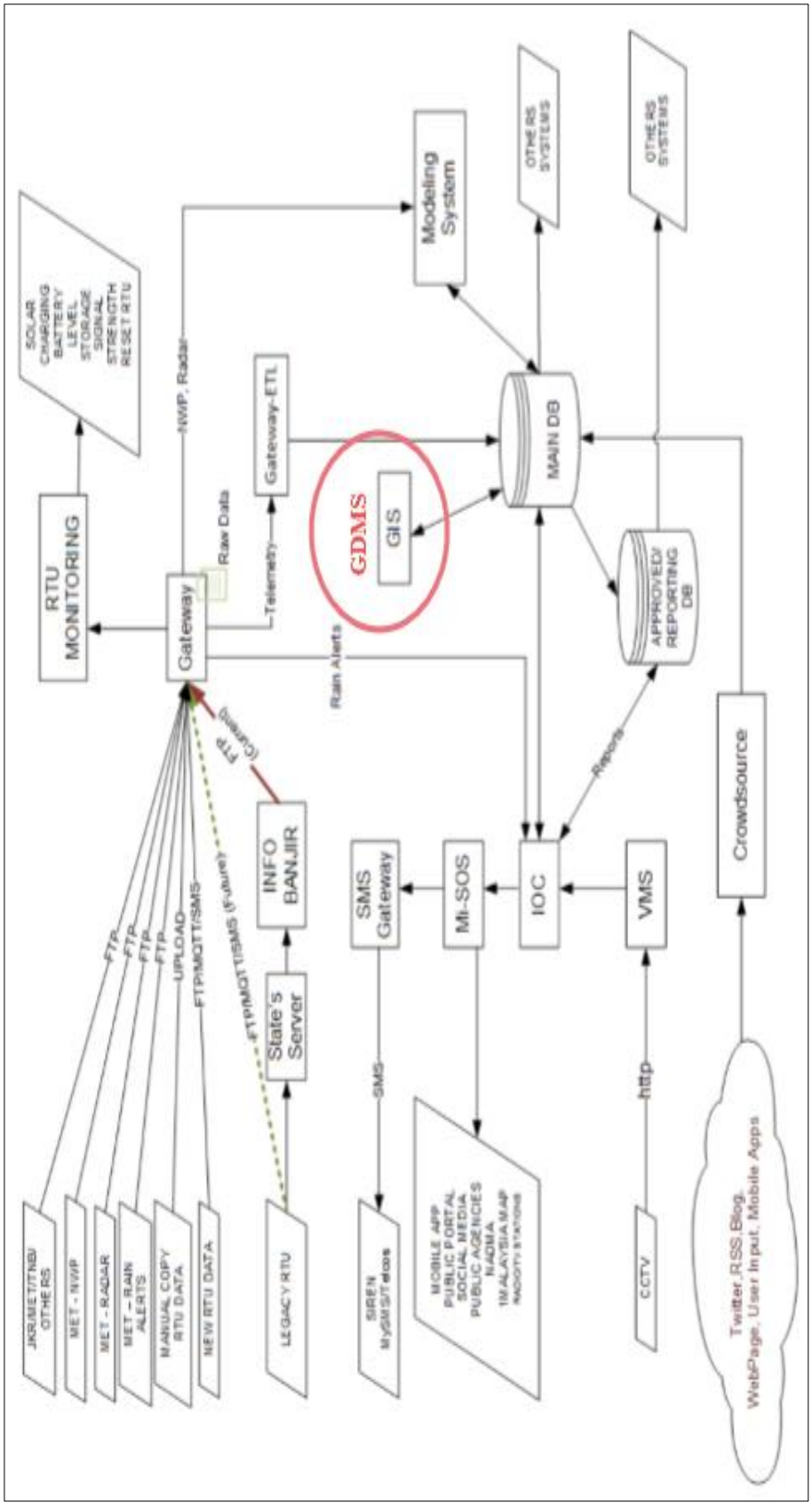


**Figure 1.3** Overall Flow Data Diagram National Flood Forecast And Warning Information Management System

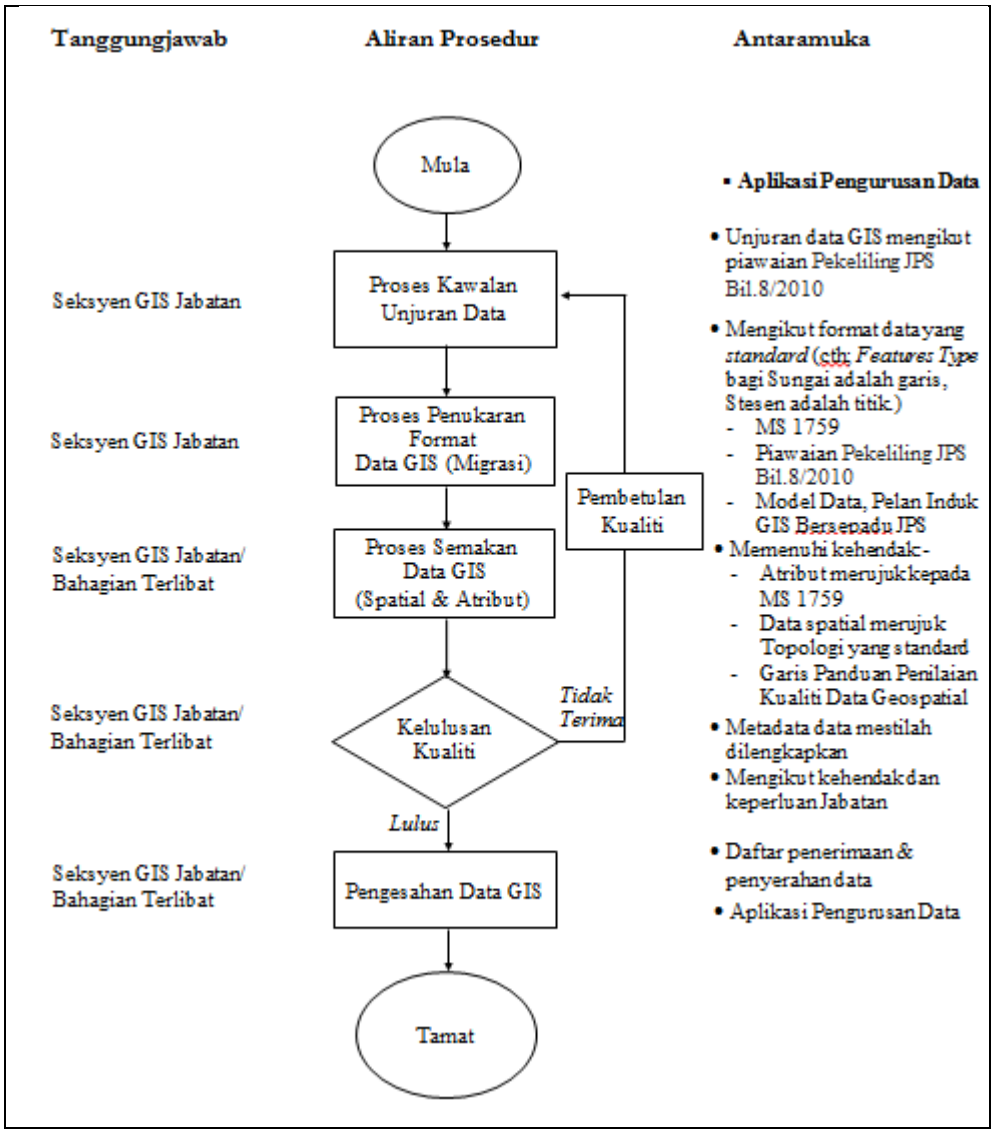


**Figure 1.4** Existing SOP for spatial data quality assessment

(*Pelan Induk Sistem Maklumat Geografi Bersepadu JPS*,2014)

Lack of knowledge on GIS Technology always a crucial issue in DFG's among staff. Only a few engineers equipped with latest GIS Technology knowledge's and it will cause them a lot of time to perform spatial data quality assessment. As an example when an Assistant Engineer need more times of inspection work for checking topology error manually as the topology rules are needed to be defined earlier by assessing the spatial data quality.

Therefore, the purpose of this project is to define a standard process flow for spatial data quality assessment for DID. The significance of having defined Process

Flow for Spatial Data Quality Assessment on Topology Data is as guidance to the engineers and IT personal to review data provided from the supplier and GIS system fulfil the GDC standard by DID Malaysia.

## 1.5 Project Objective

This project will embark the following objectives:

i. To study the spatial data quality assessment as defined in *Garis Panduan Penilaian Kualiti Data Geospatial* by Malaysian Centre for Geospatial Data Infrastructure (MaCGDI)
ii. To identify the topology rules for spatial data quality assessment for SPDG.
iii. To design the process flow for spatial data quality assessment on topology data output from SPDG
iv. To perform spatial data quality assessment on output data from SPDG

## 1.6 Project Scope

Below is the identified project scope for this project:

i. The development of Spatial Data Quality Assessment Process Flow for DID will based from the guidelines on of Spatial Data Quality Assessment by MaCGDI
ii. Spatial data quality assessment in SPDG will be focusing on topology data such as line, point and polygon

## 1.7 Importance of Project

This project was inspired by the improper spatial data quality assessment scenario at DID in order to analyse the accuracy of spatial data received from vendor or other agencies. Hence, a Spatial Data Quality Process was developed based on *Garis Panduan Penilaian Kualiti Data Geospatial* by MaCGDI with some modification to meet DID needs. From the project, it will be able to:-

i. Propose Spatial Data Quality Assessment Process Flow for DID based on Guidelines for Geospatial Data Quality Assessment by MaCGDI
ii. Improve standard of procedure for Spatial Data Quality Assessment for reviewing spatial data from supplier.

## 1.8 Software Engineering Documentation

Throughout the implementation of this project, the software engineering documentations that will be delivered are:

i. Geospatial Data Quality Assessment Guideline Studies – Studies on the best practice for Spatial Data Quality Assessment.
ii. Spatial Data Quality Assessment Procedure – Briefly explains the procedure to perform spatial data quality assessment.
iii. Spatial Data Quality Assessment Process Flow – Illustrate process involves in assessment.
iv. Procedure on Inspection Topology Rules – Steps involved to inspect the topology error based defined topology rules.

## 1.9 Project Schedule

The duration of the project is within six (6) months. The details of the activities are outlined in a Gantt Chart as in Figure 1.5.

## 1.10 Chapter Summary

The discussion in the chapter begins with project overview on GIS Technologies and the important of GIS Data to government agencies in Malaysia. This chapter also explained briefly about DID background and its core business. From the identified project background and problem statement, the project objective and the project scope is defined regarding the spatial data quality assessment. Later the importance of the project is described and the project deliverables is identified. Project schedule is later depicted in the figure which shows the activities and the timeline for each of the activities. The following chapter will describe on the literature review related with the implementation of this project. The topic covers is regarding the data quality and assessment, GIS and spatial data quality and assessment. This project is regarding the development of spatial data quality assessment focusing on topology data at DID in order to analyze the accuracy of spatial data received from vendor or other agencies. In this project, the spatial data quality assessment process flow for DID will be developed based the guidelines from MaCGDI.

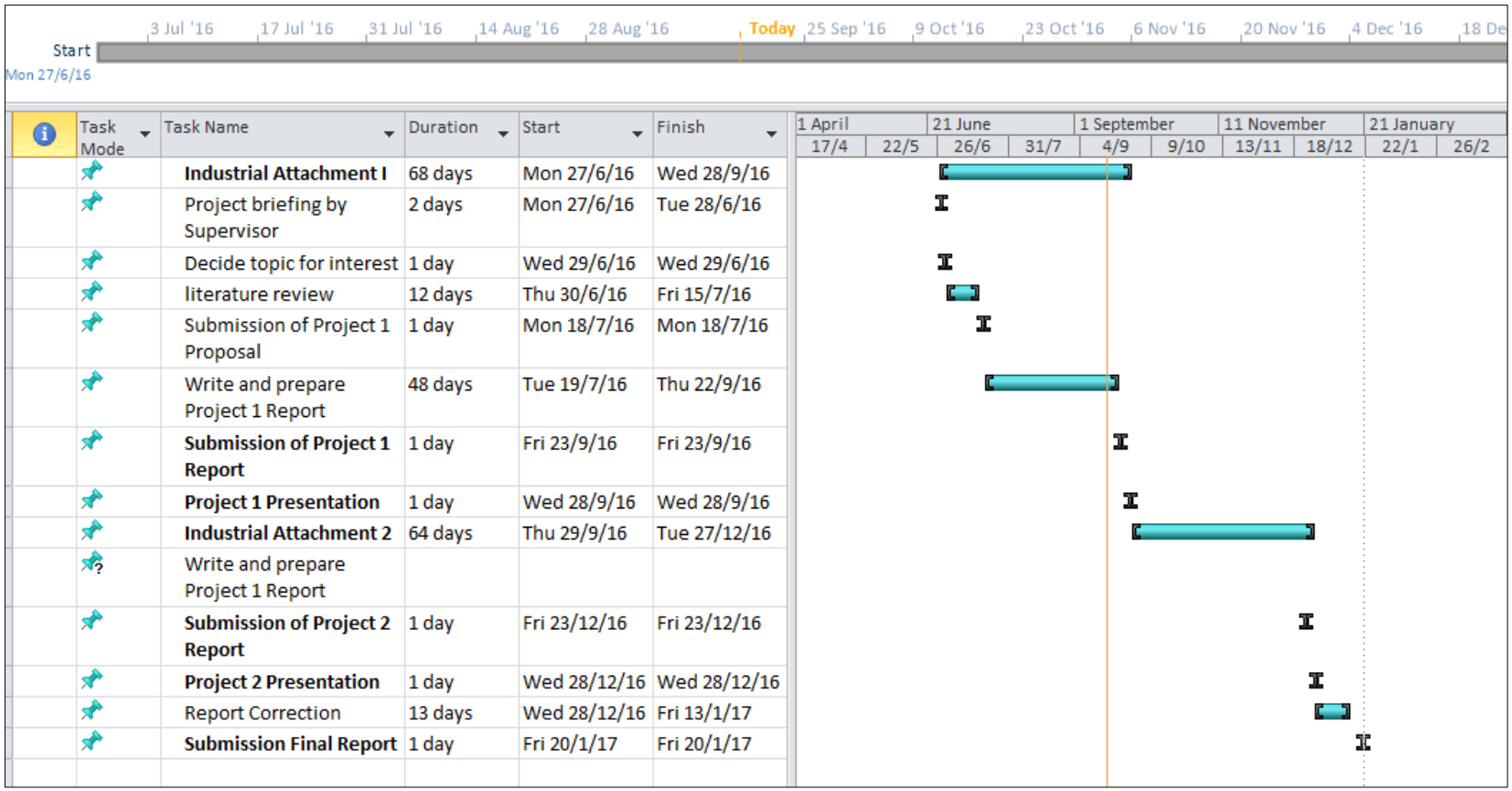


**Figure 1.5** Planning & schedule industrial attachment I & II

## REFERENCES


Ajay Jangra, Gurbaj Singh, Jasbir Singh and Rajesh Verma (2011). EXPLORING TESTING STRATEGIES. *International Journal of Information Technology and Knowledge Management*. 4(1) [January-June 2011]

Anna Brendle (2002, November 20). Police Use GIS in D.C.–Area Sniper Case and More. National Geographic News, Retrieved November 08, 2016, from http://news.nationalgeographic.com

Banque, E. (1997). Topology in Geographical Information Systems. Maps Journal. Ankara, Turkey. 118: 65-74.

Crosby, P. B. (1992) Quality is Free, McGraw-Hill, New York

Department of Irrigation & Drainage 2016, River Register Basin Phase I, Available from: < http://www.water.gov.my/home/17?date=2016-05-01&lang=my>. [09 March 2011].

Gene Rhodes, Halliburton, *The Optimum Framework for Managing E&P GIS Data*, (2013). Landmark Software & Services.

Hetzel, William C. (1988) *The Complete Guide to Software Testing.* (2nd ed) Wellesley, Mass.: QED Information Sciences.

Howell, K.E., (2013). Introduction to the Philosophy of Methodology, London: Sage Pub.

IEEE, "IEEE Standard 610.12-1990, IEEE Standard Glossary of Software Engineering Terminology," 1990

Jakob Furrer, 2010, *Automated Regression Testing of Open Geospatial Web Services.* Software Engineering Master's Thesis. University of Amsterdam, Netherlands 8[ December 2010]

Jovanovic and Irena,"Software Testing Methods and Techniques," May 26,2008.

Kualitatim. Inc 2013, Testing A GIS Application: Challenges And Perspectives. Available from: Kualitatim. Inc .[September 2012]

Malaysian Centre for Geospatial Data Infrastructure (MaCGDI) (2010). *Garis Panduan Penilaian Kualiti Data Geospatial* (1st. ed.) Putrajaya : MaCGDI

Mohd. Ehmer Khan (2010). Different Forms of Software Testing Techniques for Finding Errors. *IJCSI International Journal of Computer Science Issues*. 7(3)

Myers, Glenford , J. (1979) *The art of software testing.* New York: Wiley

Natalia Andrienko, Gennady Andrienko, Hans Voss, Fatima Bernardo, Joana Hipolito, and Ursula Kretchmer (2002) Testing The Usability of Interactive

Nicholas, D. 2013, GIS Tips – What is topology and why do I need it?. 20 August 2013. The Spatial: Blog. Available from: <https://dragons8mycat.com>. [20 August 2013].

Ondieki, C. M., & Murimi, S. K. (2009). Applications of Geographic Information Systems. In Environmental Monitoring (Vol. 2, pp. 314-340).

Ron Stanfield, (2010). A Local Government Perspective of Spatial Data Management. ArcNews, p.5

Rosmadi Fauzi (2015) Issues, Challenges And Prospects Of Application And Implementation Of Geographic Information System In Malaysia: An observation. *Malaysian Journal of Society and Space* 11(2), 118 – 127.

S.M.K Quadri and Sheikh Umar Farooq, (2010) Software Testing-Goals,Principles and Limitations. International Journal of Computer Applications, 6(9)

Sawant, A.A., Bari, P.H. and Chawan, P.M., (2012). Software testing techniques and strategies. International Journal of Engineering Research and Applications (IJERA), 2(3), pp.980-986.

Teknologi KBSE Sdn. Bhd. (2016) *Laporan Kajian Keperluan Pengguna Bagi Sistem Pengurusan Data Geospatial*. (1st. ed.) Shah Alam : MaCGDI

United States Environmental Protection Agency (2000, July). *Guidance for Data Quality Assessment. Retrieved From the Environmental Protection Agency website: https://www.epa.gov/sites/production/files/2015-06/documents*

# APPENDIX A

GIS Data Quality Assessment Procedures