AN INTEGRATED FRAMEWORK FOR
BUSINESS INTELLIGENCE SYSTEMS:
BUSINESS INTELLIGENCE PRODUCT MAP

CHEE CHIN HOONG

MASTER OF COMPUTER SCIENCE

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY
UNIVERSITI TUNKU ABDUL RAHMAN
OCT 2012

AN INTEGRATED FRAMEWORK FOR BUSINESS INTELLIGENCE SYSTEMS: BUSINESS INTELLIGENCE PRODUCT MAP

By

CHEE CHIN HOONG

A thesis submitted to the
Faculty of Information & Communication Technology,
Universiti Tunku Abdul Rahman,
in partial fulfillment of the requirements for the degree of
Master of Computer Science
Oct 2012

ABSTRACT

AN INTEGRATED FRAMEWORK FOR BUSINESS INTELLIGENCE SYSTEMS: BUSINESS INTELLIGENCE PRODUCT MAP

Chee Chin Hoong

As data is becoming more important in this information age, Business Intelligence (BI) has been adopted by many organizations to gain technological advantage over their competitors. To provide management users with actionable information (so-called BI product) for decision-making, BI incorporates various concepts, methods and technologies such as data warehousing, Extract-Transform-Load, Online Analytical Processing, reporting and data mining tools for users to analyze the data. However, current BI systems lack the modeling for business management and information manufacturing and this causes users not able to fully understand the BI product. Modeling the business management workflow and information manufacturing chain will enable users to implement Business Process Modeling (BPM) and Total Data Quality Management (TDQM). This research presents an integrated framework known as the Business Intelligence Product Map (BIP-MAP). The salient modeling and management techniques from Business Process Modeling Notation (BPMN) and Information Product Map (IP-MAP) are integrated in a hierarchical fashion for users to visualize the relevant business processes and information processes of a BI product. The issues we have studied include evaluating the efficacy of integrating these two very different types of processes and implementing data quality for the two layers. For the completeness and accuracy data quality dimensions, we have investigated how to perform automatic weighting of the data attributes to compute the two parameters so that more accurate values of data quality are presented to the management users. For the timeliness data quality dimension, to ensure consistencies between the two layers, we have proposed a novel approach to propagate the timing parameters automatically from the information process layer to the business process layer. The framework has been evaluated using two real-case studies. Both subjective evaluations confirm the effectiveness of the proposed method, BIP-MAP, over BPMN and IP-MAP.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank God for His blessings in completing my research. Then, I would like to thank my supervisor, Dr. Tan Hung Khoon, for his guidance and advice throughout this research. In addition, I would also like to thank my co-supervisor, Mr. Tan Teik Boon, who has also given me advice and guidance.

Apart from this, I would like to thank the two organizations who have participated in the survey and interview in order to evaluate the usability of the framework. Their feedback is indeed very helpful for this research.

Finally, I would like to thank my wife, Joanne Yew, for all her love, support and prayer. I am also grateful to my family for their encouragement and support during the journey of this research.

APPROVAL SHEET

This thesis entitled "**AN INTEGRATED FRAMEWORK FOR BUSINESS INTELLIGENCE SYSTEMS: BUSINESS INTELLIGENCE PRODUCT MAP"** was prepared by CHEE CHIN HOONG and submitted as partial fulfillment of the requirements for the degree of Master of Computer Science at Universiti Tunku Abdul Rahman.

Approved by:

____
(Dr. TAN HUNG KHOON)
Date:…………………..
Supervisor
Department of Computer Science
Faculty of Information & Communication Technology
Universiti Tunku Abdul Rahman

____
(Mr. TAN TEIK BOON)
Date:…………………..
Co-Supervisor
Department of Computer Science
Faculty of Information & Communication Technology
Universiti Tunku Abdul Rahman

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY

UNIVERSITI TUNKU ABDUL RAHMAN

Date: ____________________

SUBMISSION OF THESIS

It is hereby certified that **CHEE CHIN HOONG** (ID No: **09UIM09186**) has completed this thesis entitled "AN INTEGRATED FRAMEWORK FOR BUSINESS INTELLIGENCE SYSTEMS: BUSINESS INTELLIGENCE PRODUCT MAP" under the supervision of Dr. Tan Hung Khoon (Supervisor) from the Department of Computer Science, Faculty of Information & Communication Technology, and Mr. Tan Teik Boon (Co-Supervisor) from the Department of Computer Science, Faculty of Information & Communication Technology.

I understand that the University will upload softcopy of my thesis in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

____

(CHEE CHIN HOONG)

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UTAR or other institutions.

____
(CHEE CHIN HOONG)

Date ________________________

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure		Page
4.4	BIP-MAP Layer 1. This layer is constructed based on BPMN. The BI products generated at BIP-MAP Layer 2 are summarized onto this layer so that users are able to gain an overview of all the BI products that are available in the entire business process.	62
4.5	BIP-MAP Layer 2. This layer is constructed based on IP-MAP. Process B1 in Figure 4.4 is mapped to components I1, I2 and I3 where the student pre-registration report is generated by process P9 from data warehouse STO2 for the HODs.	62
4.6	BI Dashboard. The BI products are presented to the users in various forms of dashboards where the users can drill down at different parts to retrieve more detailed information.	66
4.7	Drilled Down Data of BI Product. Users are able to view detailed data of the BI product by clicking on different portions of the graph or chart.	66
4.8	Retrieve Data Module. Users are allowed to conveniently retrieve all the detailed data that is related to a business process when it is necessary.	67
4.9	Metadata for Data Validation. The main objective of this metadata is to define the conditions that need to be fulfilled by the data and indicate whether the data is compulsory to be entered by the users.	70

Figure		Page

Figure		Page

Figure		Page

LIST OF ABBREVIATIONS

BI	Business Intelligence
BIP-MAP	Business Intelligence Product Map
BPM	Business Process Modeling
BPMN	Business Process Modeling Notation
BUC	Business Use Cases
CPN	Coloured Petri-Net
EPC	Event Process Chains
ERP	Enterprise Resource Planning
ETL	Extract, Transform and Load Process
IDEF	Integrated Definition for Function Modeling
IMS	Information Manufacturing System
IP-MAP	Information Product Map
IP-UML	Information Product – Unified Modeling Language
PERT	Project Evaluation & Review Technique
RAD	Role Activity Diagram
TDQM	Total Data Quality Management
UML	Unified Modeling Language

CHAPTER 1

INTRODUCTION

1.1 Overview of Business Intelligence Systems

In recent years, Business Intelligence (BI) has emerged as one of the top ten spending priorities for many Chief Information Officers [11] and the worldwide BI software revenue surpassed US$10 billion dollars in the year 2010 alone [12]. According to [13], BI is "*a broad category of technologies, applications, and processes used for gathering, storing, accessing, and analyzing data to help its users make better decisions*". In other words, BI is a system that allows business users to leverage the disparate data sources for making informed business decisions [14]. It is becoming more popular nowadays because of its capability to extract and present critical information in a timely manner for managers to make well-informed and the best business decision.

BI is an extension of some traditional information systems like Transaction Processing System (TPS), Management Information System (MIS), and Decision Support System (DSS) [15]. Each type of these information systems plays a different role in providing support to the organizations. TPS focuses on implementing the routine activities of an enterprise by processing large amount of transactions [16]. Some of the examples of TPS are the point of sales system, inventory control system and employee payroll system. MIS converts data into a meaningful aggregated form that management users can use for business analysis [15]. For example, a MIS can transform the transaction-level data of sales and inventory to produce reports in a summarized manner for the management users to analyze the business performance of their organization. DSS provides managerial decision support for the managers and executives to implement the job of planning in their companies [16]. For instance, a DSS is able to provide some answers to the questions that are specified by the management users in order to solve a business problem.

However, these systems are still not able to meet the evolving business needs of enterprises that demand the integration of immense, different, dispersed and heterogenic data [17]. Most enterprises are still facing a great challenge in processing and analyzing their data because data is scattered at various departments of the organization [18]. The source, reliability, accuracy, and timeliness of data are essential to good decision making and successful outcomes [19], and the visualization of business process and information quality that form the basis of the managerial reports become imperative for decision makers. As a result, the focus on transaction-oriented systems in the early years of information management is starting to shift towards decision-oriented systems nowadays [20] and BI has emerged as an integrated type of information system in order to provide additional capabilities that can facilitate the integration of disparate corporate data to support the business decision making [21].

Figure 1.1 shows a typical platform of the BI system. In order to enable users to extract intelligence from the data to support their business decision making, data from various sources need to be loaded into a data warehouse through the data transformation process. Then, the data that is consolidated in the data warehouse can be loaded into the Online Analytical Processing (OLAP) tools to perform multi-dimensional data analysis or extracted with the reporting tools to generate managerial reports for business analysis and decision making.

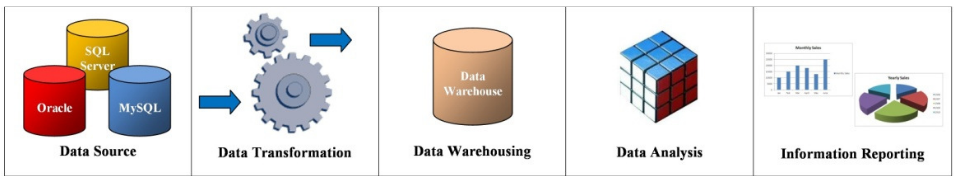


Figure 1.1: An Overview of the Business Intelligence Platform [22]

1.2 Problem Statement

In today's business environment where the operation of organization grows bigger and becomes more complex, one major challenge is how to process and extract useful information from huge volumes of data, such as customer information, transaction records, and the market and economic information. This is where a BI system is able to provide the competitive edge over others. However, one weakness of current BI systems is that data is treated too technically without considering the 'soft' aspects of the information, for example, how and why the data is generated, and how reliable the data is. Decision making is a complex process which does not only involve hard data such as business trend and sales volume but also other considerations such as the information system infrastructure, human resources, and standard operating procedures of the company. For example, any changes in the infrastructure may affect the daily operations of the business which in turn affect how data is generated, and the reliability of the BI process itself. Changes in business environment may force the company to expedite a particular business process to ensure the shortest time to market for a product.

In particular, the current BI technology has neglected the modeling of some important underlying processes that may affect the data and it does not have any mechanism to keep track of the changes for both process and data. In [20], Thomas H. Davenport proposed a bottom-up look at the information flow so that managers can identify how the *data* is manipulated throughout the *processes* in the entire organization. He introduced a four-step procedure for linking decisions and information which includes (a) identifying the decisions that are important for successfully executing the company's strategy, (b) providing information required to make the decisions, (c) recognizing the roles of different users in decision making, and (d) ensuring employees to use information for making better decisions. Although the importance of linking decisions and information has been identified and steps have been proposed for linking them, no framework has been produced to model the relevant processes and data for business decision making. As a result, the managers and business analysts are still unable to possess a complete view of the information for important business decisions. This thesis addresses these shortcomings by incorporating all the different elements of decision making so that BI systems can be robust towards the shifting needs of the company.

1.3 Research Objective

This thesis investigates the following two research issues for a BI system along this direction:

How to incorporate the elements of business planning and management into a BI system? Both experts in the fields of business and information technology have agreed that understanding the business processes is very important for developing a successful system [23] because the changes in business environment also affect the systems or applications that support its operations [24]. We investigate how to handle these issues from two perspectives. The first is in terms of the operating procedures practiced by the company (hereafter described as business processes). It addresses questions such as "what are the sequence of steps needed to perform a particular task or activity?" and "who is responsible?". Business Process Modeling Notation (BPMN) [3] is one of the modeling techniques that enables users to visualize the workflow of activities and identify the appropriate order and dependencies among them. The second takes on a more data-oriented view which specifies the steps how data is being managed in the process (hereafter described as information processes). The latter view is more technical but they are critical in the modern business environment which relies on data for decision making. It addresses questions such as "who provides the data?", "has it been verified?", "is the data up-to-date?". Information Product Map (IP-MAP) [7] is one of the modeling techniques that enables users to visualize the construction steps of an information product. One major challenge in the research is how to merge these two different models into a single system. Usually, the linking between the business and information processes is not a one-to-one mapping. In fact, different set of business processes could be linked to a set of information processes that is overlapping with one another. As BPMN and IP-MAP is integrated together, the issue of consistency for timeliness may arise when users try to perform time management for the business and information process layers. The timing parameters of the business and information processes may not be consistent in the two layers if there is no proper coordination. Chapter 4 discusses about the integration of BPMN and IP-MAP and the second part of Chapter 5 discusses about an algorithm that deals with the issue of consistency for timeliness between the two layers.

How to ensure the integrity of the data presented in a BI system? In a BI environment, data or information is the most critical element that needs to be managed well because most of the business decisions are made by the management users based on the data provided in the BI products. Since data is considered to be a very important asset of any organization, it should be viewed as a product in order to solve the problem of poor data quality [25]. When measuring the data quality problem in IP-MAP, different attributes have different impact towards the data integrity at different stages. For example, the attribute 'CustomerName' may not be as important as 'CustomerSalary' for deciding to raise the credit limit of a customer in a banking scenario. In IP-MAP [7], users have to either assign weights to the data attributes manually or treat them as equally important. However, this is not desirable as different attributes clearly has different degree of relevance in a decision making process. Apart from this, inconsistent weights may be assigned to the data attributes because different users may have different opinions towards the importance of a data attribute in the decision making process. The first part of Chapter 5 discusses about an algorithm that implements automatic weight assignment to address these issues.

1.4 Research Contribution

One major problem in incorporating BPMN and IP-MAP is how to incorporate these two very different models. The first contribution of this thesis is **a new framework that integrates the salient modeling and management techniques of BPMN and IP-MAP**. In our model, the information processes are employed to provide further breakdown the data processing steps associated with a business process. As such, the relationship between the business and information process is not a simple one-to-one mapping. In fact, some business processes may even share overlapping information processes to generate different BI products. Therefore, we propose a hierarchical model that implements a one-to-many mapping between the business processes in the first layer and the information processes in the second. The business processes will be modeled through BPMN [3] while the information processes will be modeled through IP-MAP [7].

The advantage of having a hierarchical model is that different levels of details are systematically segregated into different layers (from high-level BI products and business processes to the underlying information processes). For decision making, management users can easily identify all the BI products that are available to be utilized for business decision making. Apart from this, they are able to drill down into the detailed data processing steps that are involved in each business process in order to identify how the data is being captured, validated, processed, stored, transformed and generated throughout the organization. In addition, a comprehensive set of metadata that is included at the information process layer helps the technical users to understand some business considerations behind certain data and enables the business users to get an insight into the logic behind the information generated to them [26]. Thus, the integration of the proposed framework enables users to have a high-level understanding of all the business operations.

Since poor data quality will badly affect the usability of a data warehouse [27], the second contribution of this thesis is **enhancing the implementation of data quality for BI systems**. First, we propose a novel approach of automatic weight assignment for the data attributes based on the *Term Frequency and Inverse Document Frequency* (tf-idf) method [28] to help users identify the attributes that will severely impact the data quality of a BI product. The current methods of *Uniform Weighting* and *User Assigned Weighting* in IP-MAP are not practical to be used for identifying the data quality of a BI product due to several reasons. Uniform Weighting considers each data attribute as equally important but clearly different attributes have different degree of impact towards decision making. User Assigned Weighting requires users to manually assign the weight for each data attribute but this becomes infeasible when the number of attributes or data entries become large. This is because it will be time consuming and inconvenient for the users to enter the weight of the data attribute one by one. In the proposed data weighting method, the weight of each data attribute is determined automatically based on their usage in the BI Dashboard. With this approach, a consistent weight is always generated for each data attribute. Therefore, this approach of data weighting helps to compute the data quality dimensions of completeness and accuracy for a BI product in a more realistic manner.

Second, we propose an algorithm that enforces consistent time management between the business and information process layers. Users are only required to input the timing parameters at the information process layer for the business processes that are mapped to a set of information processes. The total duration time of the information processes will be automatically propagated up to the relevant process at the business process layer. As the duration time is generated successfully for each business process that is mapped to a set of information processes, the Project Evaluation & Review Technique (PERT) algorithm can be executed to calculate the timeliness details like early finish time, late finish time, slack time and critical path at both layers. With the implementation of this feature, consistencies for the propagation of timing parameters between the business and information process layers are achieved and users are able to estimate the time required to accomplish a task or to generate a report at different phases of a process.

1.5 Thesis Organization

The remainder of this thesis is organized as follows:

- **Chapter 2 Literature Review**: This chapter contains important information that is relevant to this research. It introduces the concept of Business Intelligence, different types of Business Process Modeling and Total Data Quality Management techniques, and Metadata.
- **Chapter 3 Research Methodology**: This chapter describes Design Science Research as the method that is applied to conduct this research and discusses the method used to evaluate the proposed framework.
- **Chapter 4 Integration of BPMN and IP-MAP**: This chapter discusses about the integration of the proposed framework using the case study of Subject Pre-Registration at Faculty of Information and Communication Technology in Universiti Tunku Abdul Rahman.
- **Chapter 5 Implementation of Data Quality**: This chapter discusses about the algorithms of the proposed framework for automatic weight assignment of the data attributes and time management between the two layers of business and information processes.
- **Chapter 6 Conclusion**: This chapter concludes the research by providing a summary about the contributions and discusses some enhancements for future research.

1.6 Research Publication

Before implementing the actual work of this research, a state-of-the-art review for BI systems is conducted to identify its applications in various industries and the findings of the study is published in the Symposium on Progress in Information & Communication Technology (SPICT 2009) [29]. In Chapter 4, the work on integrating BPMN and IP-MAP to construct the proposed framework is published in the 15th International Business Information Management Association (IBIMA 2010) Conference on Knowledge Management and Innovation [30]. After enhancing the proposed framework with additional features, the work of integrating BPMN and IP-MAP is published in the 22nd Australasian Conference on Information Systems (ACIS 2011) [31].

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This research has proposed a new framework for Business Intelligence (BI) by integrating the salient modeling and management techniques of Business Process Modeling (BPM) and Total Data Quality Management (TDQM). The previous work on BI, BPM and TDQM are discussed in Section 2.2, Section 2.3 and Section 2.4 respectively. The implementation of metadata for TDQM has been further improved in this research and the related work for metadata adoption in BI is discussed in Section 2.5.

2.2 Business Intelligence

Many organizations are facing the problem of information overload because of not implementing the appropriate systems or tools for capturing, organizing and utilizing information [32]. Business Intelligence is a technology that helps organizations to solve this problem and enables employees to extract useful information from the huge volumes of data for business analysis and decision making. The term “Business Intelligence” (BI) is popularized by the Gartner Research group in 1989 [14] and it has been defined in many ways by different researchers or practitioners of BI. Generally, BI is defined as a set of concepts, methods, and technologies used to turn data into information and knowledge that is useful for the business analysis and decision making process of an organization [33]. It should not be confused with the field of Artificial

Intelligence (AI) and Data Mining (DM). Although AI and DM techniques can be incorporated for data analysis component or as an expert system in a BI system, BI is a broader field, encompassing both technical aspects such as information system architecture, database management, as well as non-technical aspects such as business planning and management, operations research, financial reporting and others. Apart from this, BI can also be defined as the process of collecting sufficient right information with the right manner at the right time, and providing the right results to the right people for the purpose of decision making, so that it can yield true benefits for the business strategies and operations of an organization [18]. In addition, the Data Warehousing Institute defines BI as the processes, technologies, and tools that can be used to transform data into information, information into knowledge, and knowledge into plans that provide intelligence to the decision making process of business operations [34].

According to some surveys, BI applications has been placed on the top-most spending priority list of many Chief Information Officers nowadays [11, 35]. This is because BI is a prerequisite for an organization to compete with others in the marketplace [13] by helping organizations to increase profitability, decrease costs, improve customer relationship management, and decrease risks for their business [34]. In addition, BI systems are able to assist the business analysts in understanding their customers purchasing behaviour, identifying marketing opportunities, managing the progress of projects, and optimizing inventory control [36]. In fact, BI should be used by *all* employees in the organizations with the implementation of tools that have simple user interface, systems that support access control for different users, and digital documentation which provides detailed information to the users [37]. Some of the organizations that are implementing BI to capture, analyze and take appropriate actions upon their data are Wal-Mart, Harrah's, Marriott, and Capital One [38].

2.2.1 The Business Intelligence Platform

Figure 1.1 shows an overview of the BI platform [22]. Data is available in many sources throughout an organization and it may be stored in different types of databases or files. In order to implement a successful BI environment, the data stored in various sources of an organization needs to be transferred into a centralized data repository called the *Data Warehouse*. The reason of implementing a data warehouse is to present organized data for users to construct their business questions in an easy manner as they perform their business analysis and decision making process. For that reason, the *Extract, Transform and Load* (ETL) process will be executed to cleanse and validate the data before it is transferred from the data sources into the data warehouse. Then, the data can be loaded from the data warehouse for business analysis using the *Online Analytical Processing* (OLAP) and *BI Reporting* tools.

Among the many types of BI reporting tools used in various fields of business and industry like transportation, banking, health care, retail, manufacturing, and pharmaceuticals [29, 39, 40], BI Dashboard is considered as one of the important applications for providing comprehensive BI products to managers and business analysts on a single computer screen so that information can be viewed easily [41]. The information in a BI Dashboard is extracted from a data warehouse and processed into various forms of graphs or charts like the bar graph, line graph, pie chart or a combination of graphs. By using the BI Dashboard, managers and business analysts may help their organizations to save cost and gain more profit if a right decision is made based on the available historical and forecasted data.

Even though the numbers in a BI Dashboard can provide managers and business analysts with the detailed figures for sales and profit or any other information, it may not be sufficient to assist them in making informed decisions. For example, if the sales of a product are very low, users may not have any idea about the cause of this situation since they do not understand how information in the BI Dashboard is being constructed. The current BI solutions do not take into consideration the underlying factors that affect the construction of the BI products. In order to implement a successful BI solution, the management users are required to understand how the data is captured and stored throughout an organization for addressing the complicated issues of business decision making [20].

Apart from this, any changes in the business environment or operating procedures may result in a different requirement of the BI products. This is because different set of information is required to perform analysis for different processes in business decision making. A gap will appear if the data and processes are analyzed separately using different BI and BPM tools [42]. Therefore, we proposed to implement a BI framework based on some relevant processes that will produce a great impact towards the construction of the BI products available in a BI Dashboard. These processes can be categorized as the business process and the information process. Section 2.3 provides a few types of Business Process Modeling (BPM) techniques for modeling the business process while Section 2.4 provides some techniques of Total Data Quality Management (TDQM) for modeling the information process. Each modeling technique has its own strengths and weaknesses.

2.3 Modeling of Business Processes

Recently, business processes become the main focus of many organizations for supporting the analysis and design of their information systems [43]. A business process [44] includes a set of activities that are executed in a synchronized manner to achieve some business objectives. It is performed in an organization and it may interact with the business processes from other organizations. Business Process Modeling (BPM) [23] is the technique used to represent the processes of an enterprise in a systematic form, so that the current processes may be analyzed and improved in future. Typically, managers and business analysts are responsible to improve the efficiency and quality of a business process by referring to the BPM diagram because it provides them a clear picture of the business process workflow. In order to implement improvement for the business processes, managers and business analysts should ensure quality for the execution of six business process flows, including activity flow, information flow, resource flow, cost flow, cash flow, and profit flow [45]. By conducting a proper modeling for the business processes, an organization will be able to perform better in its business operations. This is because the employees can easily understand the standard operating procedures of a business process when each step of the process is modeled systematically.

Many types of BPM techniques have been proposed by different researchers for the fields of business and information technology. Some examples of the BPM techniques [23, 46, 47] are the Role Activity Diagram (RAD), Event Process Chains (EPC), Business Use Cases (BUC), Unified Modeling Language (UML), Integrated Definition for Function Modeling (IDEF), Coloured Petri Net (CPN), and Business Process Modeling Notation (BPMN). Each BPM technique plays a specific role in modeling the business processes of an organization. The strengths and weaknesses of each BPM technique are discussed from Section 2.3.1 to Section 2.3.6.

Table 2.1 shows a comparison of the BPM techniques for its process flow, data flow and users interaction. All the evaluated BPM techniques are having a clearly defined process flow except for the BUC technique. This is because each activity in the BUC is individually linked to the users that are involved and it will create a complicated structure for the process flow when the number of activity increases. Presenting a well organized business process flow to the users is important because it helps them to conduct a business operation in a successful manner when all the procedures are clearly defined. Not all the evaluated BPM techniques focus into modeling the flow of data. The BPM techniques that do not highlight the flow of data in its modeling are the RAD, EPC, BUC and the Activity Diagram in UML. Without a proper modeling of the data flow, a BPM technique will not be able to present an accurate view of its process flow because different input and output of an activity will affect the implementation of a business process. For users interaction, the BPM techniques that present a clear and organized view are the RAD, CPN and BPMN. This is because the activities to be conducted by different users are categorized into the appropriate sections. Indicating a clear interaction between the users in a business process will help to identify the right person to provide a solution when problems occur to a business operation.

Table 2.1: Comparison of BPM Techniques

Model	Process Flow	Data Flow	Users Interaction	Understandability	Flexibility
RAD [2]	• Clearly modeled. • Activities to be performed by different users or systems are grouped appropriately.	• Not modeled.	• Very organized. • Activities to be conducted by the users are categorized appropriately.	• Easily understood by all levels of users.	• Easily updated without much technical knowledge.
EPC [8]	• Clearly modeled. • Technical operators like AND, OR and XOR are used for modeling.	• Not modeled.	• Not visible. • Activities are not modeled according to the roles of users.	• Easily understood by all levels of users.	• Easily updated without much technical knowledge.
BUC [4]	• Not clearly modeled. • Each activity is individually linked to the users that are involved.	• Not modeled.	• Not organized. • Activities are linked individually to the users instead of being grouped according to the roles of users.	• Easily understood by all levels of users.	• Easily updated without much technical knowledge.
UML [1]	• Clearly modeled in the Activity Diagram. • The start and end of each activity is defined properly.	• Not modeled in the Activity Diagram.	• Not visible in the Activity Diagram.	• May not be easily understood by the business users.	• Requires the knowledge in UML modeling to update the diagram.

Table 2.1 continued: Comparison of BPM Techniques

Model	Process Flow	Data Flow	Users Interaction	Understandability	Flexibility
IDEF [5]	• Clearly modeled with the appropriate control. • The input and output of data is specified for each activity.	• Input and output of each activity are clearly defined.	• Not visible. • Activities are not modeled according to the roles of users.	• Easily understood by all levels of users.	• Easily updated without much technical knowledge.
CPN [6]	• Clearly modeled. • Various states of a system are indicated.	• Input and output of each activity are clearly defined.	• Very organized. • Activities to be conducted by the users are categorized appropriately.	• May not be easily understood by the business users.	• Requires mathematical knowledge to update the diagram.
BPMN [3]	• Clearly modeled. • Activities to be performed by different users are grouped into separate sections.	• Message of communication between the users is presented clearly.	• Very organized. • Activities to be conducted by the users are categorized appropriately.	• Easily understood by all levels of users.	• Easily updated without much technical knowledge.

2.3.1 Role Activity Diagram

Role Activity Diagram (RAD) [2] models a process based on the responsibility of individual participants within the process and indicates the interactions between them. Figure 2.1 shows an example of the RAD for implementing a project [47]. The activities to be conducted by users or executed by systems within a process are categorized into different groups in the RAD. This enables the business analyst to improve and revise the activities without influencing the entire process. As the activities are grouped together for the appropriate roles, the managers are able to visualize the business process in a clearer manner and make decision easily for improving the business operations. However, RAD is unlikely to be decomposed and this will cause too many details to be included at one level of the diagram [23]. Apart from this, data is not represented in the RAD and this prevents the business analyst from identifying the exchange of data between the interactions of users.

2.3.2 Event Process Chains

Event Process Chains (EPC) [8] is constructed with three fundamental components namely event, function, and connector. Figure 2.2 shows an example of the EPC for a book borrowing process [47]. The event and function objects have exactly one input and one output, while the connector objects are allowed to have multiple input and one output, or vice versa. EPC is implemented in the commercial software and it is supported by many companies that provide solutions for Enterprise Resource Planning (ERP) and Business Process Reengineering (BPR). Nevertheless, EPC has unclear definitions for some of its components. For example, the propagation of a process folder at the OR and XOR connectors cannot be determined locally because whether a process folder can arrive at the other connector or not relies on the entire process flow of the EPC model [48]. Consequently, this may lead to failure in the execution of a business process. Lately, EPC has been extended to construct reference models for business processes so that standard business processes can be adopted at different organizations [49].

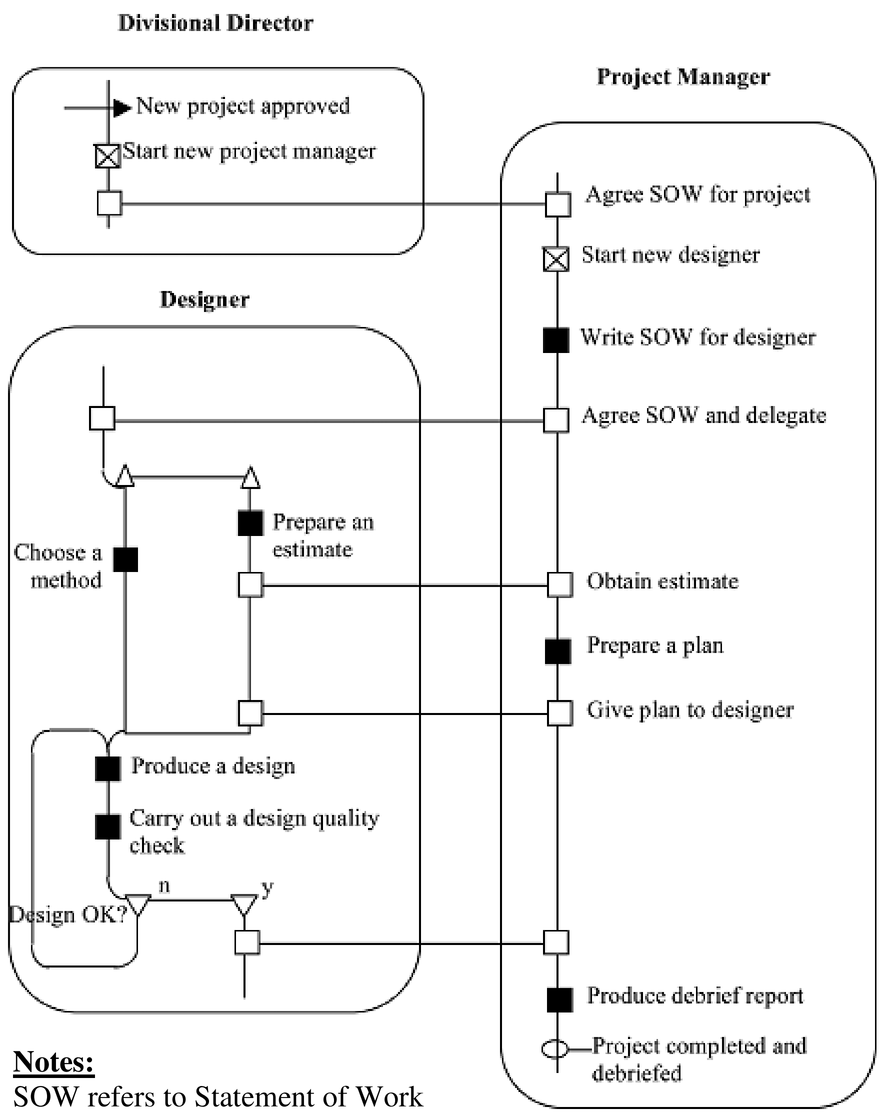


Figure 2.1: Role Activity Diagram [47]

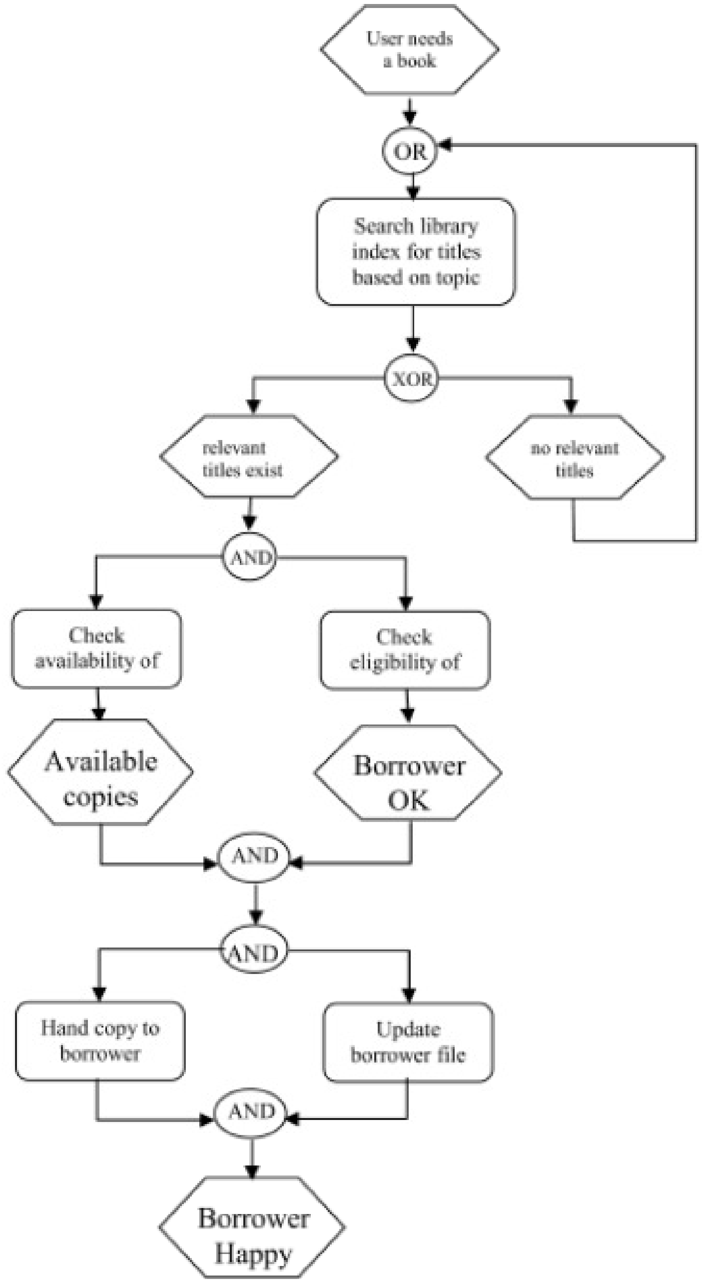


Figure 2.2: Event Process Chains [47]

2.3.3 Business Use Cases

Business Use Cases (BUC) [4] is a diagram with textual descriptions that illustrate the business processes involved at providing a service to the users. Figure 2.3 shows an example of the BUC for a course registration process [46]. Since BUC is constructed using the natural language, it is easily understood even by the non-technical users. It provides flexibility in usage because the textual descriptions can be easily updated by the business users. But this can also create problems to the business operations as the textual descriptions may be inconsistent when it is modified by different users. To avoid these problems, the textual descriptions can be refined by combining them with some simple graphical representations that provide users a clearer comprehension of the business process.

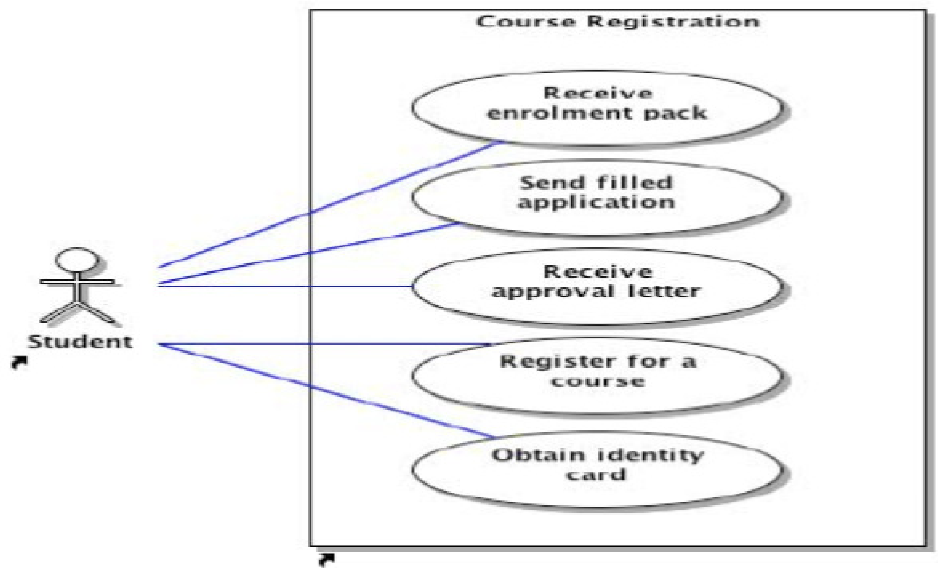


Figure 2.3: Business Use Cases [46]

2.3.4 Unified Modeling Language

Unified Modeling Language (UML) [1] is an object-oriented modeling method used to specify, visualize, construct and document software and non-software systems [23]. Figure 2.4 shows an example of the UML Class Diagram [50]. It consists of nine different diagrams, with each diagram showing a specific aspect of the system, namely Class Diagram, Object Diagram, Statechart Diagram, Activity Diagram, Sequence Diagram, Collaboration Diagram, Use-Case Diagram, Component Diagram, and

Deployment Diagram. It is a very good modeling technique for the system development team to analyze their applications and design a proper solution that meet the system requirements [51]. A total of 13 diagrams are available in UML2 but only the Activity Diagram is suitable to be used for modeling business processes [52]. This is because the Activity Diagram provides notations to model the flow of activities in a business process, especially those involving strategic decisions [53] whereas other diagrams are mostly used for modeling object-oriented systems [52]. As a result, most of the components in UML are rarely used for BPM.

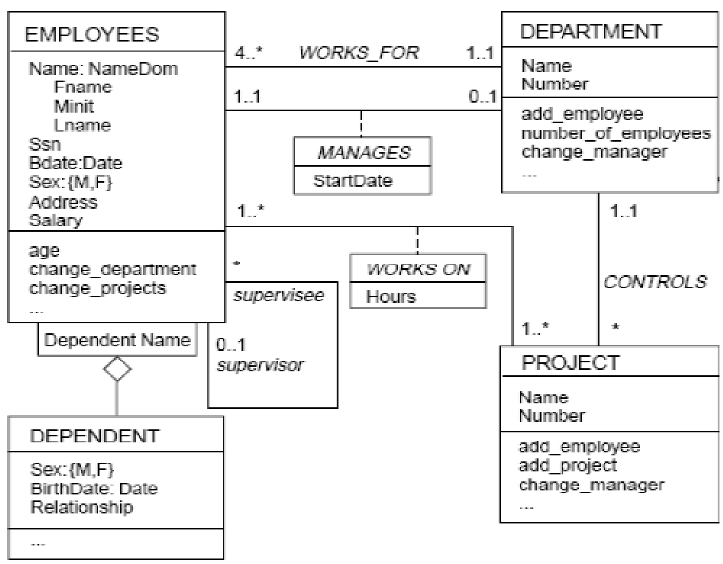


Figure 2.4: UML Class Diagram [50]

2.3.5 Integrated Definition for Function Modeling

Integrated Definition for Function Modeling (IDEF) [5] is a set of techniques used to support the business modeling needs of an organization. Figure 2.5 shows an example of the diagram for IDEF0 [23]. IDEF has been proposed in response to the needs of improving the manufacturing processes of US Air

Force in the mid-1970s and it is grouped into different methods, namely IDEF0, IDEF1, IDEF1X, IDEF2, IDEF3, IDEF4 and IDEF5. Among the various methods, IDEF0 and IDEF3 are the most relevant versions to be used for business process modeling because IDEF0 specifies the function of processes by showing the high-level activities which can be further decomposed, while IDEF3 shows different views about how processes work together within an organization. However, these two models also consist of some weaknesses. IDEF0 does not indicate the roles of a participant in the business process and too many partial diagrams are utilized to describe a business process in IDEF3.

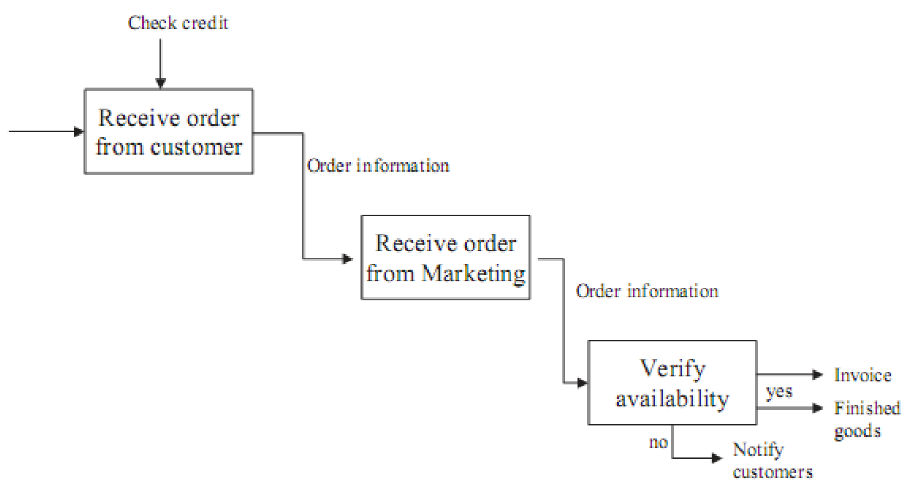


Figure 2.5: IDEF0 Diagram [23]

2.3.6 Coloured Petri Net

Coloured Petri Net (CPN) [6] is a mathematical method that is suitable to be used for modeling systems which contain a lot of processes that communicate with one another in a synchronize manner. Figure 2.6 shows an example of the diagram for CPN [23]. It is recognized as the most appropriate method to be used for describing and analyzing the communication, synchronization and resource allocation between parallel processes. Recently, the modeling technique of CPN has been enhanced with a timing function to represent the critical timing variation in a context-aware system like the collision avoidance system for vehicles that are moving on the road [54]. However, the models produced with CPN are extremely large and it is time consuming for users to construct the entire model of a business process.

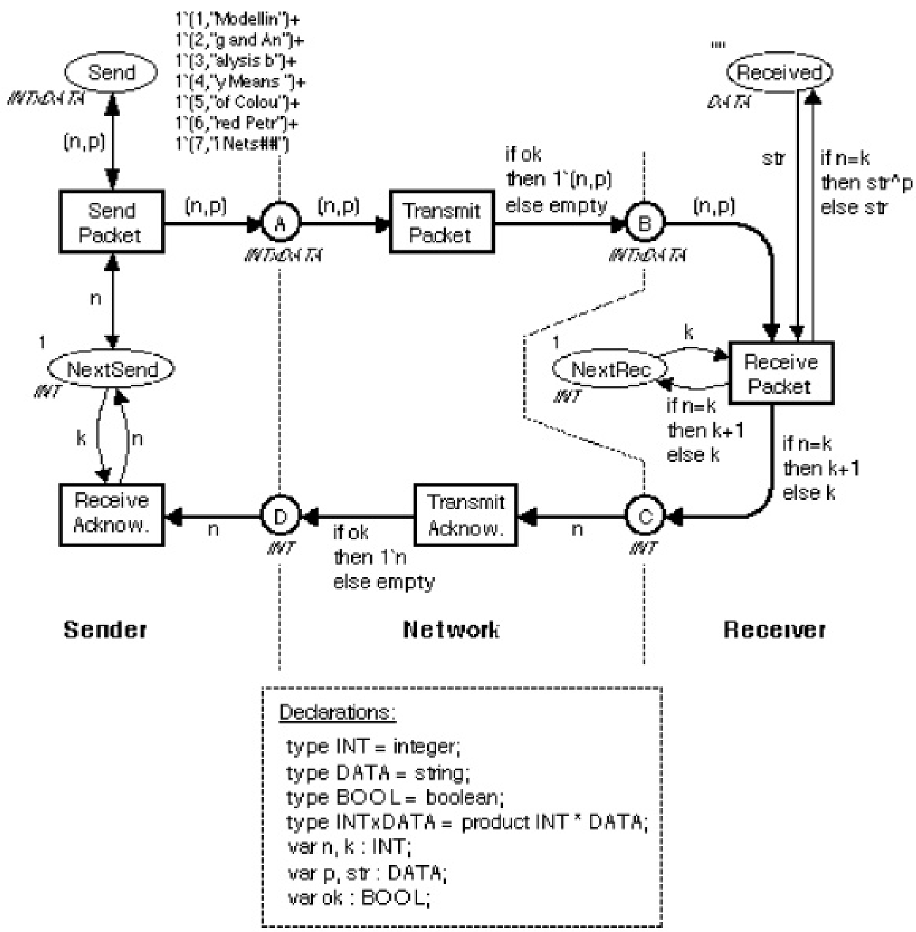


Figure 2.6: Coloured Petri Net Diagram [6]

2.3.7 Business Process Modeling Notation

Business Process Modeling Notation (BPMN) [3] is a model that is developed by the Business Process Management Initiative (BPMI) based on some flowcharting technique. Figure 2.7 shows an example of the diagram for BPMN [3]. The detailed description of each component in BPMN is provided in Section 4.2. The ultimate aim of BPMN is to provide a standardized notation that is easily understood by all business users and technical developers for the purpose of managing business processes.

BPMN is the most dominant modeling technique for business processes [55] where it is found to have a clearer structure and easier to use compared to other techniques [52]. The notations are easily understandable by all level of users, including the business analysts that model the processes to the technical developers that implement the systems for executing those processes [56]. The organized structure of BPMN is able to represent the interaction of business processes executed by different participants in a clear and simple manner. In addition, BPMN helps to define the business collaborations and transactions within and between organizations so that Business-to-Business (B2B) activities can be conducted easily [57]. However, BPMN does not model a complete flow of data to assist users in understanding how data is manipulated throughout the entire organization.

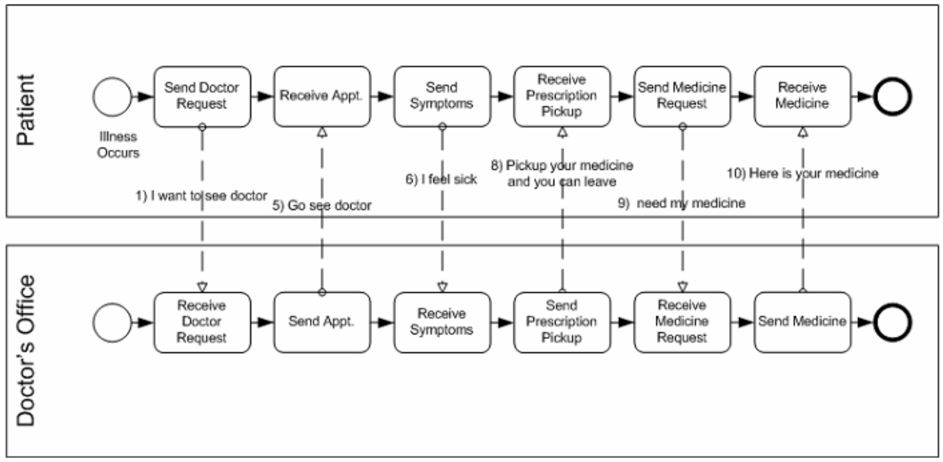


Figure 2.7: BPMN Diagram [3]

2.3.8 Summary of the BPM Techniques

Among the various types of BPM techniques, BPMN has been selected to be integrated into the proposed framework of this research because it provides a standardized notation that is easily understood by all levels of users in an organization for the purpose of managing business processes. Although most of the BPM techniques enable users to gain a clear picture of the business process workflow and some of them (IDEF, CPN and BPMN) model certain flow of data, they do not provide users the details about how data is manipulated throughout a business process. The way how data is captured, validated, processed, stored, transformed and generated is not clearly defined. Therefore, users are not able to visualize the entire information manufacturing chain that is essential to implement Total Data Quality Management (TDQM) and the ability to access meaningful data may be one of the major problems in BPM [58]. To conduct a successful BPM, it is necessary to include a detailed modeling of data into the BPM techniques so that users are able to access the relevant data of a business process and understand how data is manipulated in the entire organization.

2.4 Modeling of Information Processes

The issue of data quality is a major problem faced by many organizations because the existence of invalid data will definitely affect the results of analysis for data that is manipulated by a set of information processes [59]. An information process includes a set of system or data processing activities that are executed by different types of information systems in an organization. Total Data Quality Management (TDQM) [60] is a practice that delivers high quality information to consumers of an information product by treating data and information much like the products found in a manufacturing environment of any industry. For example, TDQM is applied to the field of health care to reduce medical errors and thus decrease the cost and improve the caring service for patients [61]. To implement TDQM, the information processes of an organization need to be modeled in a systematic manner. Information Product Map (IP-MAP) is the modeling technique that implements the concept of TDQM. It is an extended and modified version of the Information Manufacturing System (IMS) and being enhanced to become the Information Product – Unified Modeling Language (IP-UML). The strengths and weaknesses of these three modeling techniques are discussed in Section 2.4.1, Section 2.4.2 and Section 2.4.3.

Table 2.2 shows a comparison of the manufacturing for a physical product and an information product. In the manufacturing industry, physical products are produced by processing the raw materials in the assembly line. Similarly, information products are produced by processing raw data with some information systems. The major difference is that data can be used multiple times by more people and not consumed, but material can only be utilized once for manufacturing a single physical product [62].

Table 2.2: Product vs. Information Manufacturing [62]

	Product Manufacturing	**Information Manufacturing**
Input	Raw Materials	Raw Data
Process	Assembly Line	Information System
Output	Physical Products	Information Products

In order to perform a better job for any business activities, data should be managed in a proper manner so that everyone is accessing the high quality and timely information. According to [62], four roles have been identified to implement TDQM in an organization; they are the Information Suppliers, Information Manufacturers, Information Consumers and Information Product Managers. Information Suppliers are the people that gather and provide data into the information systems for generating the information product, while Information Consumers are the people that receive and use the information product in their work. The data and systems infrastructure of an information product are designed, developed and maintained by the Information Manufacturers. Last but not least, the Information Product Managers are the people that manage the entire manufacturing process of any information product in an organization.

Four steps are involved in the TDQM cycle as shown in Figure 2.8 [60]:

1. Define the Information Product
    - Information quality requirements of an information product should be assessed in a detailed manner.
2. Measurement
    - Information Products (IP) should be verified to determine its fitness to be used by consumers of an organization.
3. Analysis
    - Information quality problems and its effects should be identified at this stage.
4. Improvement
    - Methods to improve information quality should be delivered from time to time.

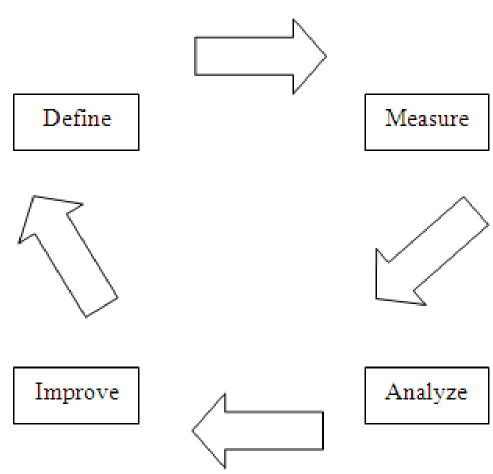


Figure 2.8: Total Data Quality Management Cycle [60]

Table 2.3: Comparison of TDQM Techniques

Model	Process Flow	Data Flow	Users Interaction	Understandability	Flexibility
IMS [10]	• Not clearly modeled. • Sequence of execution for the activities is not properly defined.	• Clearly modeled with the appropriate building blocks. • The way how data is captured, processed, verified and stored into a system is presented clearly.	• Not visible. • Activities are not modeled according to the roles of users.	• Easily understood by all levels of users.	• Easily updated without much technical knowledge.
IP-MAP [7]	• Not clearly modeled. • Sequence of execution for the activities is not properly defined.	• Further enhanced with metadata to describe the actual meaning of each data.	• Not visible. • Activities are not modeled according to the roles of users.	• Easily understood by all levels of users.	• Easily updated without much technical knowledge.
IP-UML [9]	• Not clearly modeled. • Sequence of execution for the activities is not properly defined.	• Further improved with some diagrams from the UML modeling technique.	• Very organized. • Activities to be conducted by the users are categorized appropriately.	• May not be easily understood by the business users.	• Requires the knowledge in UML modeling to update the diagram.

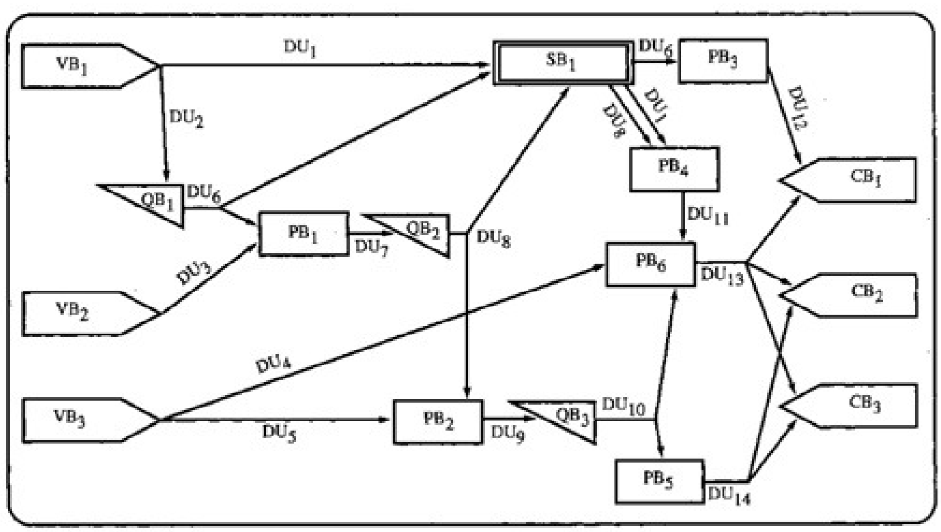
Table 2.3 shows a comparison of the TDQM techniques for its process flow, data flow and users interaction. All the evaluated TDQM techniques do not present a clear process flow because the sequence of activities is not indicated in its modeling. The way how data is being manipulated throughout an organization is clearly modeled in all the three techniques. However, IMS and IP-MAP do not present a proper interaction among the users that are involved because the activities are not modeled according to the roles of users. IP-UML presents a clear interaction among the users by categorizing the activities to be conducted by different users into the appropriate sections.

2.4.1 Information Manufacturing System

Information Manufacturing System (IMS) [10] is a modeling technique that is developed to implement information quality management by imitating the product quality management in a manufacturing environment of any industry. Figure 2.9 shows an example of the IMS [10]. The model represents a set of activities (so-called information manufacturing process) that are designed to process the raw data in order to produce the information product where its timeliness, quality, cost and value are also determined.

Five different components are used to describe the entire information manufacturing process; they are the Data Vendor Block, Processing Block, Data Storage Block, Quality Block and Customer Block. Each block plays a specific role in defining different parts of the information manufacturing process. The Data Vendor Block represents the input of raw data that comes from different sources, while the Customer Block represents the output of information to be used by the relevant users. Any process of manipulation, calculation and combination for data is represented by the Processing Block. The Quality Block represents any validation process that is implemented to enhance the quality of data before it is stored into the database or file system which is represented by the Data Storage Block.

The limitation of IMS [63] is that it focuses on calculating the quality of the final information product only. Therefore, TDQM is not fully supported by IMS where the data quality should be accessible at all phases of an information manufacturing process. However, IMS serves as a foundation for IP-MAP to be developed as a model that helps users to comprehend the construction of an information product.



VB_n	: Data Vendor Block
CB_n	: Customer Block
PB_n	: Processing Block
QB_n	: Quality Block
SB_n	: Data Storage Block
DU_n	: Data Unit

Figure 2.9: An Example of Information Manufacturing System [10]

2.4.2 Information Product Map

Information Product Map (IP-MAP) [7] is a modeling technique used to implement the concept of TDQM by representing the construction of an information product in a systematic manner. Figure 2.10 shows an example of the IP-MAP [7]. It is an extended and modified version of the IMS since IMS does not collect or provide quality-related information in its model [64]. The aim of IP-MAP is to help people comprehend, evaluate and describe how an information product is assembled [60]. By using IP-MAP, one can visualize the construction of an information product and measure its quality with suitable dimensions in order to conduct constant improvement for the data quality of an organization. The detailed description of each component in IP-MAP is provided in Section 4.3.

As data becomes an important asset to the organization, it is vital to implement TDQM in order to ensure that high quality data is utilized by the users to support their decision making. The systematical representation of the information processes enables the users to visualize how the data is being captured, validated, processed, stored and generated throughout an organization. When they are able to identify a proper workflow of the entire information manufacturing chain, it helps them to apply different methods in manipulating the data for various business analyses and decision makings. Furthermore, they can gain a deeper understanding of an information process because metadata is provided to describe the detailed steps of data processing. On the other hand, IP-MAP does not model the flow of business processes that are closely related to the information processes in a chronological manner. Its modeling is too focused on the data itself and it is detached from the objective and operating procedures in the organization. In other words, the human-factor and business-factor are missing. Consequently, when a business process is changed, users are not able to easily identify the underlying information processes that may be affected.

2.4.3 Information Product – Unified Modeling Language

Information Product – Unified Modeling Language (IP-UML) [9] is a framework that combines the modeling techniques of IP-MAP and UML for supporting the improvement of data quality in an organization. It is constructed with three models namely the Data Analysis Model, Quality Analysis Model and Quality Design Model. Figure 2.11 to Figure 2.13 show the examples of the three models [9]. The Data Analysis Model indicates the data that is available, how the data is being composed or derived in the information manufacturing process. The Quality Analysis Model specifies the requirements of quality that need to be fulfilled by each type of data while the Quality Design Model links the data and processes together for supporting data quality improvement.

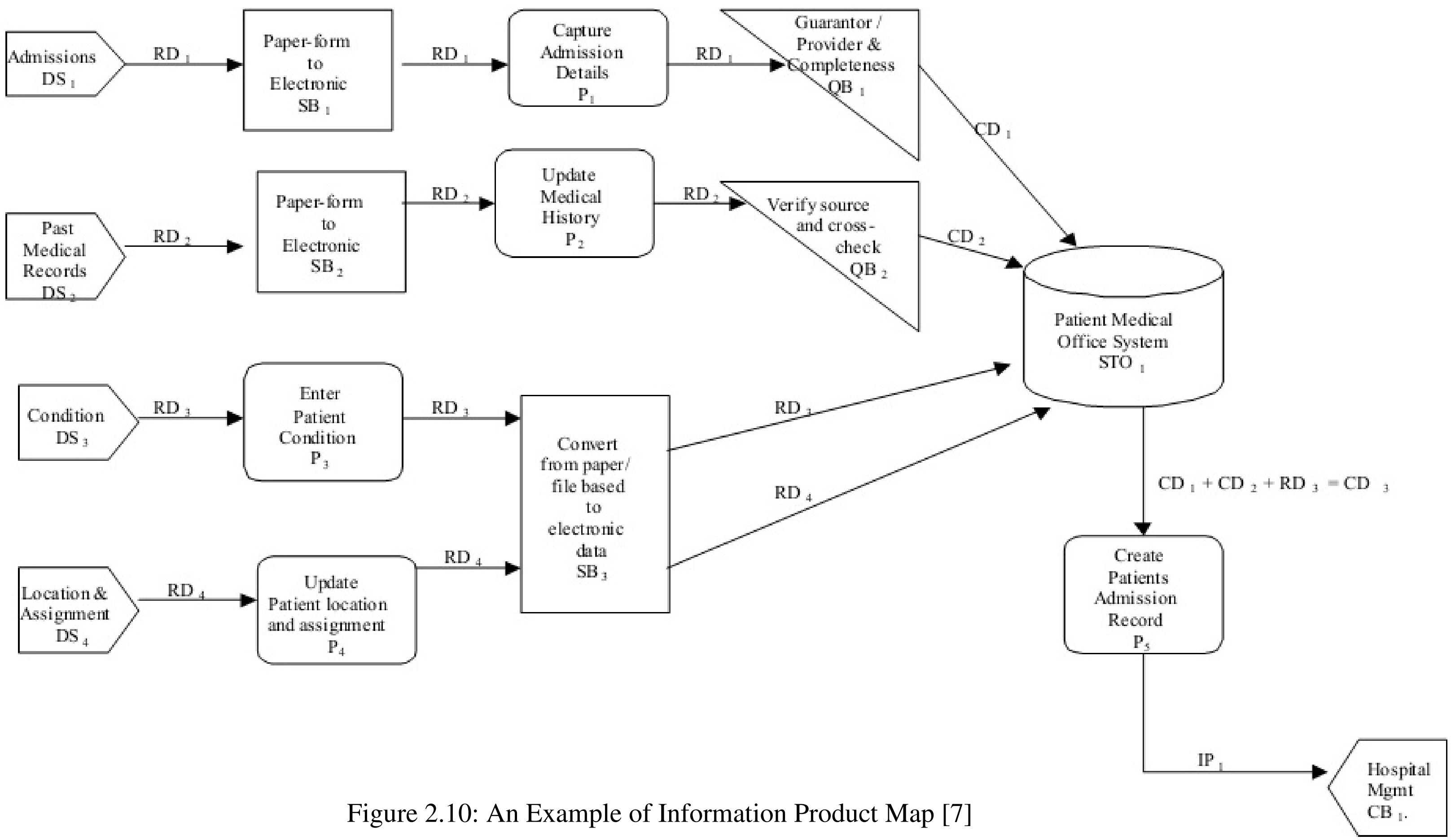


Figure 2.10: An Example of Information Product Map [7]

By integrating IP-MAP and UML into the IP-UML framework, users are able to identify the relationships and dependencies among different type of data, distinguish the quality requirements of data for different quality dimensions, and recognize the combination of data and processes that are involved for quality improvement. However, even though IP-UML has utilized the swimlanes object to separate the processes into different sections for replacing the Information System Boundary block and Business Boundary block in IP-MAP, it still does not present a standardized notation that is easily understood by all level of users compared to BPMN.

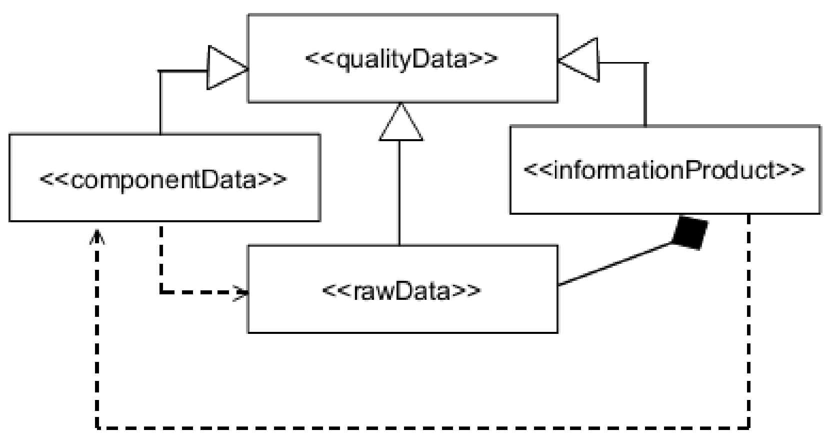


Figure 2.11: An Example of the Data Analysis Model in IP-UML [9]

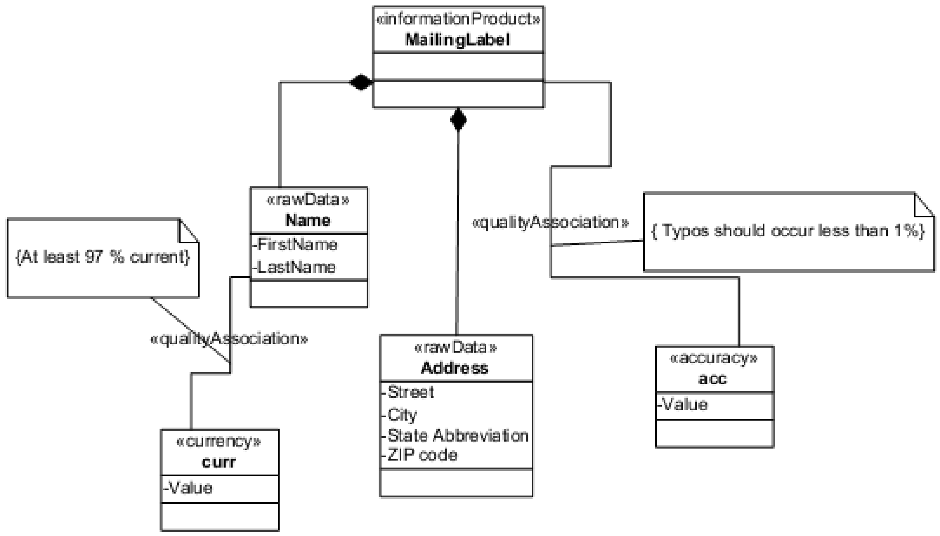


Figure 2.12: An Example of the Quality Analysis Model in IP-UML [9]

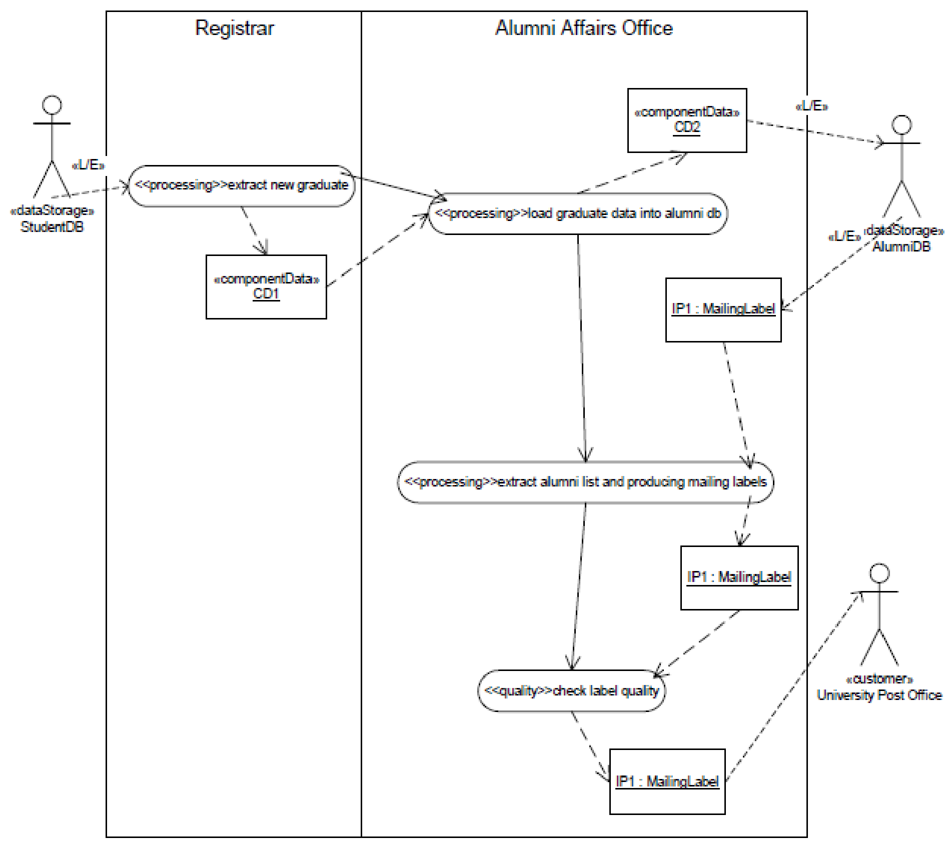


Figure 2.13: An Example of the Quality Design Model in IP-UML [9]

2.4.4 Summary of the TDQM Techniques

Among the three TDQM techniques, IP-MAP has been chosen to be integrated into the proposed framework of this research to implement TDQM because it presents a systematical representation of the processes that are involved in the manufacturing of an information product. In spite of the fact that TDQM is able to deliver high quality information to the users when all the information processes that are involved in manipulating the data is properly modeled, all the three TDQM techniques still lack the modeling for the business processes that may affect the quality of data that is captured into the systems of an organization. The change of a business process flow will indirectly cause an effect on the requirements of data that is supposed to be captured or generated. Thus, it is necessary to include the modeling techniques of BPM into TDQM so that the business processes that are closely related to a set of data can be properly modeled in order to produce high quality information to the users.

2.5 Metadata for Business Intelligence

Metadata is generally defined as the data about data [65] and its role is to provide details or explanation for some data in a data warehouse. The challenge of BI is to extract real intelligence from the huge amount of data that is entering the system in our daily business transactions and BI is more than just capturing and extracting data around [38]. In order to extract useful data for business analysis, metadata is definitely required to be made available in a data warehouse so that the management users are able to identify the real meaning of data before utilizing it for any business decision making. Since the importance of metadata has been emphasized by most of the enterprises that view data as a valuable asset of the organization [66], we have included metadata into the implementation of the proposed BI framework. This section discusses about the significance and implementation of metadata for the current BI solutions.

Metadata can be classified in many different ways like the technical and business metadata, back-end and front-end metadata, and metadata categorized based on its functionality [26]. The *business metadata* will provide the meaning of data to business users, and the *technical metadata* is meant for the team that is responsible to develop and administer the data warehouse. Similarly, the *back-end metadata* is used to describe the data storage and processing, and the *front-end metadata* is used to describe the method of information delivery and usage. Metadata that is grouped *based on its functionality* is used to describe data in various aspects like infrastructure, model, process, quality, delivery, and administration.

Metadata provides a few significant functionalities [65] towards the implementation of BI in an organization. First, it helps managers and business analysts to understand the data before they can analyze the information that has been retrieved with the appropriate queries to make decision for their business. Apart from this, metadata provides users the details for building a robust data warehouse because developers need to know the data structures, source-to-target mappings and rules of data transformation as they construct the data extraction, transformation and loading (ETL) processes. In addition, metadata is also important for administering the data warehouse since most of the data warehouses are complex and huge in size.

Figure 2.14 shows the application of metadata in a BI environment. The perceived quality of data is affected by the applications and users of data because the same data may be viewed at different perspectives by different people in different situations [64]. If the meaning of data is understood wrongly by users during the decision making process, it may cause severe consequences to the business of an organization. Therefore, metadata should be implemented properly at various parts of a BI system in order to ensure that data is utilized correctly for any case of business decision making.

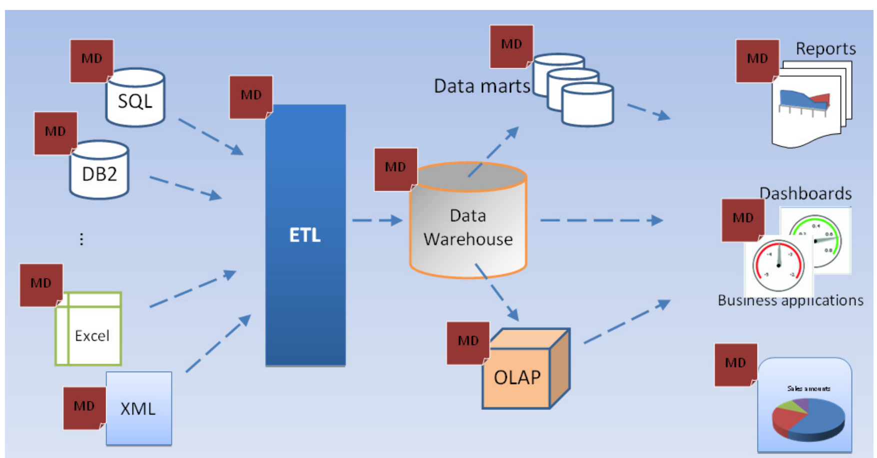


Figure 2.14: Metadata in a BI Environment [67]. Metadata should be implemented properly at various phases of a BI environment like the data source, data transformation process, data warehouse, data analysis, and information reporting stages.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

Selecting a suitable methodology is important for producing quality results in a research. In this chapter, we discuss about the research methodology that we have adopted in pursuing the research in this thesis. There exists different types of methodologies [68] such as Math Modeling, Experimental Simulation, Laboratory Experiment, Free Simulation, Field Experiment, Adaptive Experiment, Field Study, Group Feedback Analysis, Opinion Research, Participative Research, Case Study, Archival Research, and Philosophical Research. We have selected the Design Science Research approach [69] to perform a systematic investigation on integrating the two models of Business Process Modeling Notation (BPMN) and Information Product Map (IP-MAP).

3.2 Design Science Research

Design Science Research [69] is a methodology that involves the design of artifacts and analysis of its functionalities in order to improve the performance of an information system. For example, the artifacts include algorithms, programming languages, software or system interfaces, system analysis and design methodologies, and process or data models. In general, a Design Science Research method consists of a few phases like artifacts design, development, analysis and assessment. The ultimate aim of this method is to acquire the knowledge and understanding of a problem and solution through the development and application of some new and innovative artifacts. Table 3.1 provides seven guidelines that are derived from the fundamental principle of Design Science Research for researchers to conduct their research using this methodology in a successful manner.

Table 3.1: Design Science Research Guidelines [70]

Guideline	**Description**
1. Design as an Artifact	Workable artifact must be developed in the form of an instantiation, a method, a model, or a construct.
2. Problem Relevance	The purpose of this methodology is to construct technical solutions for the appropriate business problems.
3. Design Evaluation	The usefulness and effectiveness of a design artifact need to be thoroughly evaluated.
4. Research Contributions	Contributions in the areas of artifact design must be clearly provided and provable.
5. Research Rigor	Precise methods should be applied for the development and evaluation of the design artifact.
6. Design as a Search Process	Research should be conducted iteratively in order to search for a best design for the problems in information systems.
7. Communication of Research	The technical users and business users must be able to comprehend the work of research.

3.3 Research, Development and Evaluation

This research follows the Design Science Research approach and comprises the development and evaluation of an artifact. The artifact that has been designed and developed is an integrated framework for Business Intelligence

(BI) systems (Guideline 1 – Design as an Artifact). To verify the usefulness of the artifact, purposive sampling is applied where the appropriate evaluators for the artifact are specifically selected instead of randomly [71]. Rather than conducting a statistical sampling, replication logic is performed by having the framework implemented for two organizations. It is implemented for the *Subject Pre-Registration* of *Faculty of Information and Communication Technology* (FICT) at *Universiti Tunku Abdul Rahman* (UTAR) and a few processes in an online job recruitment firm.

The purpose of this research is to address the problem of users who do not fully understand the manufacturing process of a BI product and the actual meaning of data that is available to them, and to provide a better support for the users in their decision making (Guideline 2 – Problem Relevance). In order to achieve this, the salient modeling and management techniques from BPMN and IP-MAP are integrated in a hierarchical fashion for users to visualize the relevant business processes and information processes of a BI product (Guideline 5 – Research Rigor).

The management users that have evaluated the framework are listed in Table 3.2. The framework is evaluated by them via a number of in-depth and semi-structured interviews (Guideline 3 – Design Evaluation). Each interview session lasted for about 1 to 1.5 hours. Two sets of survey questionnaires have been designed to gather some feedbacks from the management users in order to evaluate the functionalities of the framework. The survey questionnaires are attached in Appendix A and Appendix B.

To establish the quality of this research, the criteria that have been investigated are *construct validity*, *internal validity*, *external validity*, and *reliability* [72]. In order to implement construct validity, the survey questionnaires are carefully designed so that the evaluators are able to provide the relevant feedback. To achieve internal validity, the evaluators are briefed about the functionalities of the framework and given the opportunity to perform hands-on testing before answering a series of questions to evaluate its usefulness. For external validity, the framework is evaluated by the management users from two organizations of different backgrounds, with one in the academic and the other in the industry. Last but not least, to ensure reliability, the framework is evaluated by the evaluators who are well experienced in their field of management so that constructive feedback can be gathered to refine the designed artifact for future works. Such a build-and-evaluate loop will iterate several times before the final design artifact is generated (Guideline 6 – Design as a Search Process).

The major contributions of this research include the integration of business and information processes, enhancement of data quality and implementation of time management (Guideline 4 – Research Contributions). Users are able to make informed decision when the models of BPMN and IP-MAP are being integrated to construct a framework that maps the business and information processes together. The details of this contribution are described in Chapter 4. Apart from this, users are able to identify the importance of different data attributes and gain a more realistic data quality computation when the weight of each data attribute is automatically calculated by the framework.

Furthermore, users are able to perform time management at a two-layered framework where the timeliness details of each process are summarized from the information process layer to the business process layer. The details of these contributions are described in Chapter 5. Moreover, the framework will enable the technical users to understand the business considerations that underlie certain data while the business users are able to obtain insights into the logic used to generate a BI product (Guideline 7 – Communication of Research).

Table 3.2: Research Participants in the Case Organizations

Universiti Tunku Abdul Rahman	
Participant	**Designation**
1	Dean
2	Deputy Dean
3	Deputy Dean
4	Deputy Dean
5	Head of Department
6	Head of Department
7	Head of Department
8	Former Head of Department
9	Former Head of Department
10	Course Coordinator
Online Job Recruitment Firm	
Participant	**Designation**
1	Chief Executive Officer
2	Chief Information Officer
3	Chief Operations Officer
4	Chief Technology Officer
5	Regional Process Improvement Manager
6	Quality & Admin Manager
7	Management Information Systems Manager

3.4 Chapter Summary

The Design Science Research approach is selected as the methodology to conduct this research. An integrated framework for BI systems is designed and developed for two organizations in order to demonstrate the functionalities of the framework. To confirm the usefulness of the functionalities, the framework is evaluated by the management users through a few sessions of interviews where two set of survey questionnaires are prepared to gather their feedbacks. The evaluation results will serve as a guide for future research to further improve the framework.

CHAPTER 4

INTEGRATION OF BPMN AND IP-MAP

4.1 Introduction to Business and Information Processes

The major objective of Business Intelligence (BI) is to provide actionable information (so-called *BI product*) to the management users so that they can perform appropriate analysis for their decision making in the organization [41]. Examples of BI products include any kind of reporting information like sales performance, financial status, inventory control and others. BI products are usually presented to the users in various forms of graph, chart or table. Having the ability to analyze the historical and current data will enable users to identify a better solution for a complicated problem. However, a lot of processes are involved in constructing the BI products throughout the entire organization. This encompasses the processes for data entry, data validation, data processing, data transformation, data generation and any business activity or operating procedure that is related to the development of the BI products. It is crucial to identify all the relevant processes because any change in a process has an effect on the BI products. As the workflow of a process is changed, the reporting information of a related BI product may not be accurate for decision making if the procedures of constructing the data for the BI product is not updated accordingly.

Since data plays an important role for providing insight to the users, it is necessary to correctly model all the relevant processes that are involved in constructing the BI products. In general, these processes can be categorized as the *business process* and *information process*. A business process includes various business activities or operating procedures that are performed by different employees of an organization to achieve a business objective, whereas an information process includes the system or data processing activities that are executed by different types of information systems. In order to produce BI products with actionable information for decision making, it is useful to integrate the techniques of Business Process Modeling (BPM) and Total Data Quality Management (TDQM) into the BI systems of an organization. This is because BPM provides users a proper way of modeling the business processes while TDQM gives users the appropriate method of modeling the information processes and ensuring high quality data to be used for supporting their decision making. Among the various types of BPM and TDQM techniques, Business Process Modeling Notation (BPMN) and Information Product Map (IP-MAP) have been selected to develop the integrated framework. The reasons of selecting these two models have been discussed in Section 2.3.7 and Section 2.4.2.

This chapter is organized as follows. First, a brief introduction to BPMN and IP-MAP are provided in Section 4.2 and Section 4.3 respectively. Then, Section 4.4 describes the functionalities and advantages of the proposed framework in this research. Finally, Section 4.5 discusses the results of evaluation conducted by some management users for the framework.

4.2 Introduction to Business Process Modeling Notation

Business Process Modeling Notation (BPMN) [3] is a modeling technique that is based on the flowcharting method. It is enriched with notations that are specifically designed to perform Business Process Modeling (BPM) in order to manage the business processes of an organization. The model is developed using simple notations that are easily understood by different levels of users at both the business and technical areas. With BPMN, users are able to simply identify the flow of activities and the interaction among all the participants in a business process. This enables an organization to execute or improve its business operations in a quick and easy manner.

The elements of BPMN can be categorized into four basic groups like *Flow Objects*, *Connecting Objects*, *Swimlanes*, and *Artifacts*. The function of each element in the four groups is briefly described from Table 4.1 to Table 4.4. The elements in the Flow Objects group are used to describe the occurrence of events and the execution of activities in a business process. Connecting Objects group contains elements that connect all the objects together in order to indicate the flow and dependencies among the events and activities. The elements in the Swimlanes group categorize the activities to be conducted by different users into separate sections so that the interaction between all participants that are involved in a business process can be easily identified. Artifacts group contains elements that are used to include additional information about how the activities are executed in a business process. Even though BPMN consists of many notations, users are not required to understand all of them in order to construct a useful BPMN diagram [46].

Table 4.1: BPMN Flow Objects [3]

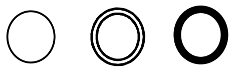
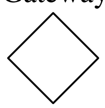
Flow Objects	Function
Event  Start End Intermediate	- Event is something that happens during the execution of a business process and it is categorized as Start, Intermediate, and End.
Activity 	- Activity is the work performed in a business process.
Gateway 	- Gateway is used to control the direction of flow for the activities in a business process.

Table 4.2: BPMN Connecting Objects [3]

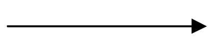
Connecting Objects	Function
Sequence Flow 	- Sequence Flow shows the order of activities being performed in a business process.
Message Flow 	- Message Flow shows the communication of messages between two different parties.
Association 	- Association is used to indicate the inputs and outputs of an activity.

Table 4.3: BPMN Swimlanes [3]

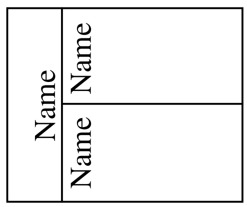
Swimlanes	Function
Pool 	- A participant in a process is represented by a pool that acts as a graphical container for separating a set of activities.
Lane 	- The sub-section of a pool is called a lane and it separates the pool, either vertically or horizontally.

Table 4.4: BPMN Artifacts [3]

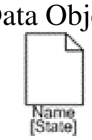
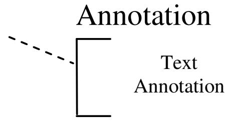
Artifacts	Function
Data Object 	- Data Object shows the data that is captured or generated by an activity.
Group 	- Group categorizes a set of activities together for the purpose of documentation or analysis.
Annotation 	- Annotation provides extra details about an event or activity.

Figure 4.1 shows an example of BPMN for an invoice approval process [73]. Three users are involved in this business process and they are the accounting department, first approver and second approver. All the processes are categorized into separate sections with the swimlanes object in order to show the user that is responsible to execute it. The start event object indicates that the process will begin when an invoice is received by the accounting

department. The entire workflow of the invoice approval process is represented in a logical manner with the activity object, sequence flow object and gateway object. The end event object indicates that the process is finished when the invoice is booked.

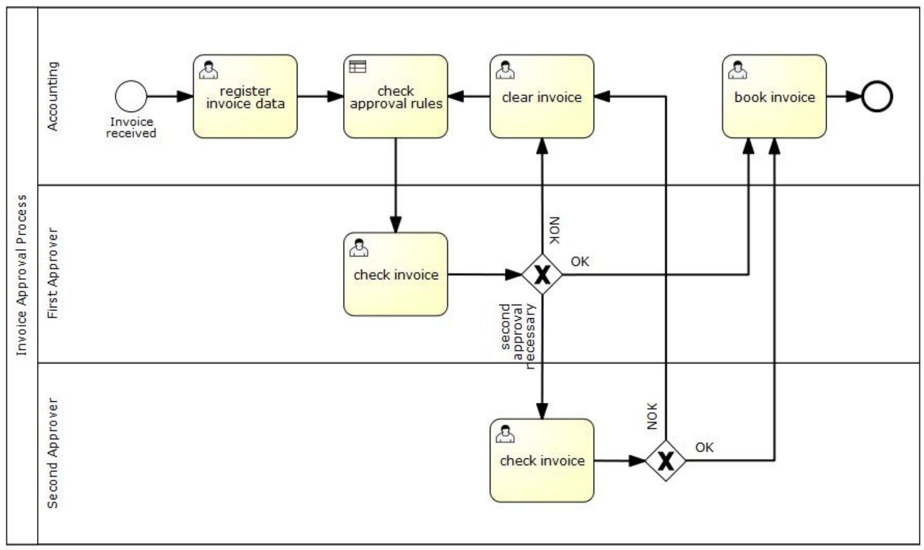


Figure 4.1: An Example of BPMN for Invoice Approval Process [73]

4.3 Introduction to Information Product Map

Information Product Map (IP-MAP) [7] is a modeling technique for representing the manufacturing processes of the information products available in an organization. In IP-MAP, information is being managed like a product in order to implement Total Data Quality Management (TDQM). With IP-MAP, users are able to visualize the entire information manufacturing chain and identify the critical parts that may create data quality problems.

The function of each building block in the IP-MAP model is briefly described in Table 4.5. The *Source* and *Customer* blocks play the role of indicating the origin and destination of data so that users are able to identify all the providers and consumers of data. The *Processing*, *Data Quality* and *Decision* blocks are used to show the processes that the data has gone through and its flow from one place to another based on different scenarios. The *Data Storage*, *Business Boundary* and *IS Boundary* blocks specify where the data is being stored or transferred throughout the organization.

Table 4.5: IP-MAP Building Blocks [60]

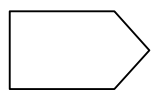
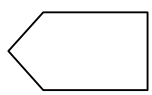
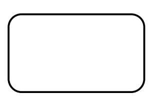
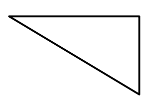
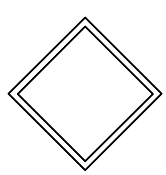
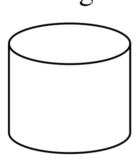
Building Blocks	**Function**
Source Block 	- Represents the source of raw data or input.
Customer Block 	- Represents the consumer of an information product.
Processing Block 	- Represents any manipulations, calculations, or combinations involving the raw input data or component data.
Data Quality Block 	- Represents the checks for data quality on data items.
Decision Block 	- Used to capture different conditions to be evaluated and direct it to the appropriate procedures.
Data Storage Block 	- Represents the capture of data in storage files or databases.

Table 4.5 continued: IP-MAP Building Blocks [60]

Building Blocks		Function
Business Boundary Block 	-	Represents instances where the raw input data or component data are "handed over" by one business (organizational) unit to another unit.
IS Boundary Block 	-	Used to reflect the changes to raw or component data as they move from one information system to another information system.

Figure 4.2 shows an example of IP-MAP for a patient's treatment history in a health care environment [7]. The Lab Radiology Reports (DS_5), Surgical Unit (DS_6), Specialist (DS_7) and Attending Physicians (DS_8) are the data providers that supply raw data to the system. The Care Providers (CB_2) are the doctors or nurses that play the role of data consumers and they receive the patient treatment history as the information product. Processes that are involved can be categorized as the data capturing process (P_6, P_7, P_8, P_9), data validation process (QB_3, QB_4) and data generation process (P_{10}). In TDQM, it is the Data Quality Block that enforces the implementation of data quality. The data validation process helps to capture quality data into the system by verifying and matching the data with the appropriate criteria before it is stored into the database (STO_2). In this example, the diagnostic results are matched with the patient records before it is being stored into the database. Apart from this, the data is transferred (BB_1) from the surgical unit to the patient care unit and it is also converted (SB_4, SB_5) from paper format to electronic form. As the information processes are represented in a systematic manner, organizations

will be able to implement TDQM. The implementation details of TDQM are provided in Section 2.4.

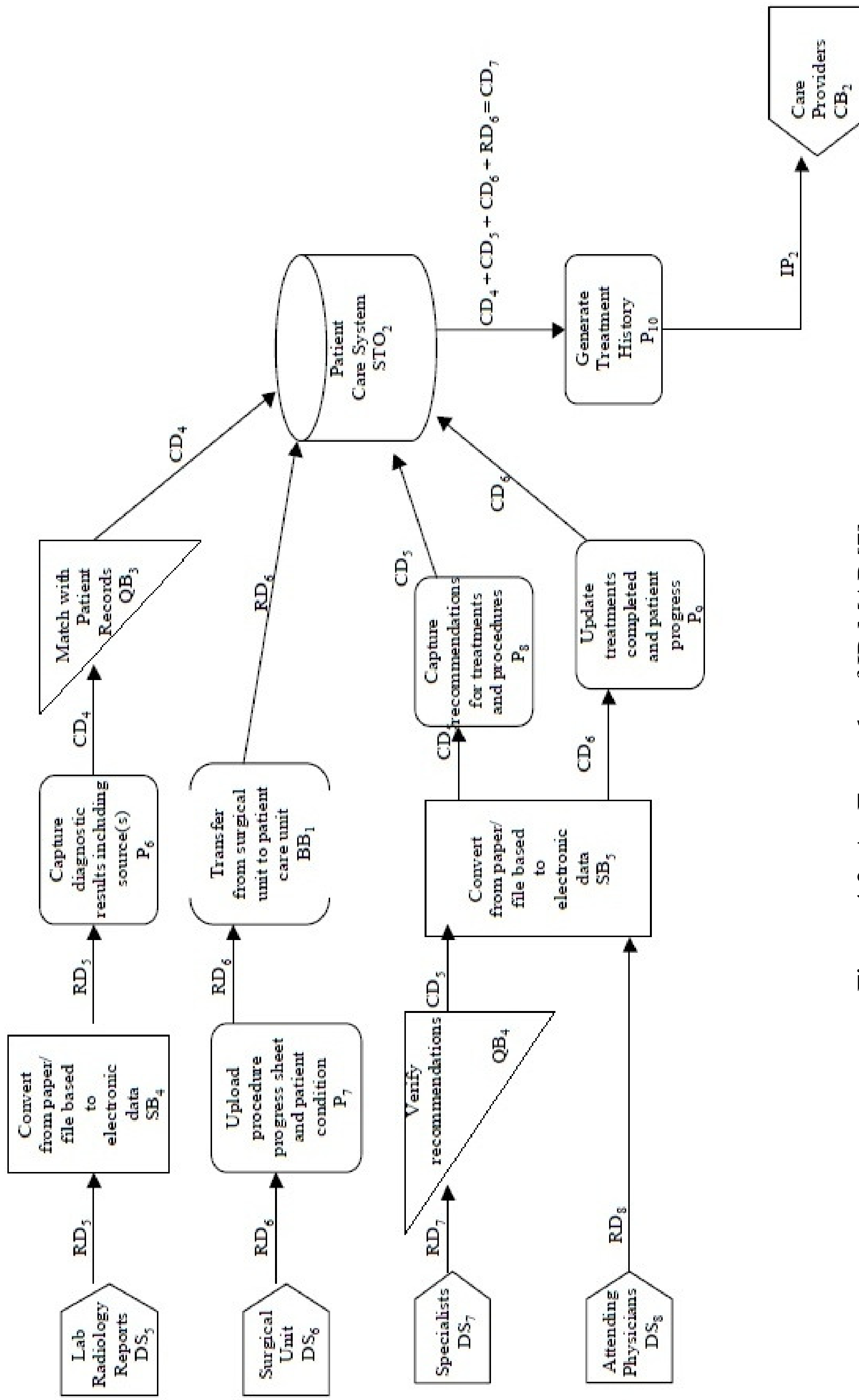


Figure 4.2: An Example of IP-MAP [7]

4.4 Proposed Framework: Business Intelligence Product Map

In this research, the significant modeling and management techniques of Business Process Modeling Notation (BPMN) and Information Product Map (IP-MAP) are being integrated to construct the Business Intelligence Product Map (BIP-MAP). BIP-MAP consists of two layers, the first layer models the business processes of an organization while the second layer models the underlying information processes that are utilized to generate the BI products within a business process. Figure 4.3 shows the hierarchical architecture of BIP-MAP. At the first layer, the management users are able to visualize the entire workflow of a business process and identify all the BI products that can provide them useful information for business analysis and decision making. Before making a decision based on the available data, they can access the second layer to visualize the information process and further understand the actual meaning of data by referring to the metadata.

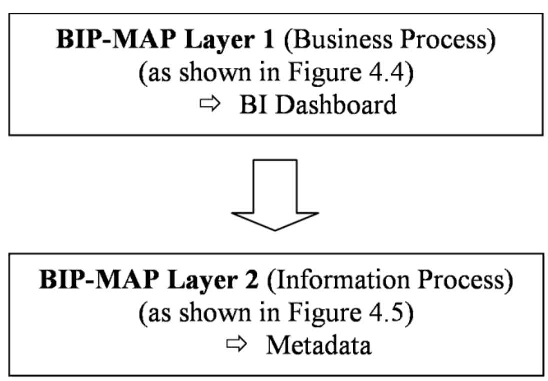


Figure 4.3: The Hierarchical Architecture of BIP-MAP

To facilitate further discussion, for the remaining of this thesis, a real example is used to demonstrate the structure and functionalities of the proposed framework. BIP-MAP is implemented for the *Subject Pre-Registration* at the

Faculty of Information and Communication Technology (FICT) in Universiti Tunku Abdul Rahman (UTAR). First, we give a brief introduction to the subject pre-registration activity. In each trimester, students are required to pre-register for the subjects they wish to study in the next trimester. The process commences as a series of activities engaged by different participants. Initially, the course structure of a programme is constructed by the Curriculum Design Committee (CDC) and reviewed by the Faculty Academic Development and Curriculum Committee (FADCC). Then, the Faculty Board (FB) will review the course structure again before the University Academic Development and Curriculum Committee (UADCC) recommends it to the Senate for approval. If the course structure is approved, the Faculty General Office (FGO) will create the course structure and enter the details of all subjects into the system. To commence the procedure of subject pre-registration, the FGO will prepare the projected student number for each subject based on the actual registration figure of the previous trimester. After that, the Head of Department (HOD) will refer to the projected student numbers to allocate classes and assign lecturers or tutors for each subject so that the students can pre-register the subjects that they prefer to study in the next trimester. If any changes are required for their pre-registered subjects, they are still allowed to add, drop or withdraw any subject in the next trimester.

4.4.1 BIP-MAP Layer 1: Business Process Modeling for BI Product

The main focus of BI is to transform data into knowledge by providing actionable information to the management users as a guideline to assist them in decision making. The actionable information is termed as a BI product in this thesis. A BI product includes any type of information that is presented to the users in different types of graph, chart or table. The information included in a BI product is constructed with a lot of raw data that is extracted from various parts of a data warehouse. Numerous tools have been developed to generate various forms of BI products to help users analyze the large volume of corporate data in a visualize manner. But the current BI technology has neglected the modeling of some important underlying processes that may affect the data in a BI product.

In today's competitive environment, it is common for the business processes of an organization to be changed from time to time. The change of business processes may cause some systems and applications to be updated as well in order to support the new business operations. As the systems and applications are changed, the data that is generated into the BI product may not be correct anymore if the procedure of constructing the data for the BI product is not updated accordingly. The current BI systems do not have any mechanism to keep track of such changes. *Therefore, the technique of Business Process Modeling (BPM) is proposed to be incorporated into a BI system* to indicate 'what' BI products are available, 'where' and 'when' they are produced and 'why' they are required by the users, especially when the business processes are changing rapidly with the business environment. The details about 'how' the BI product is generated are described in the information process layer at Section 4.4.2.

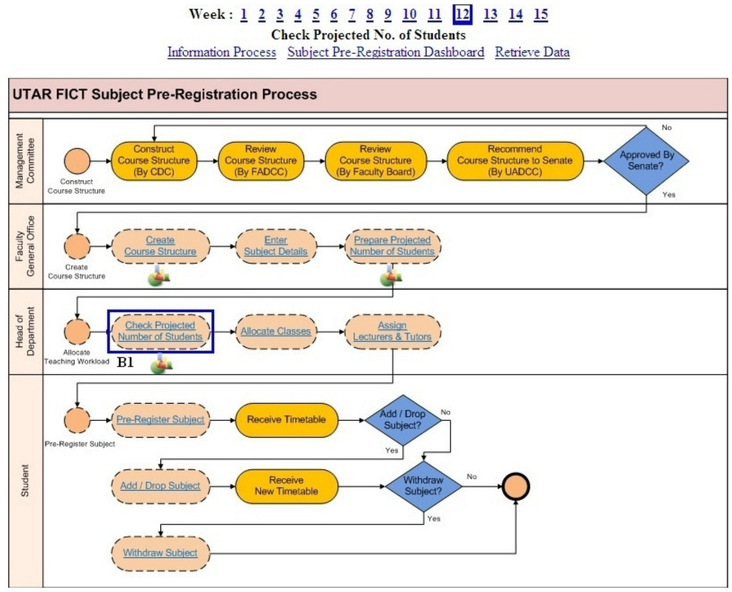


Figure 4.4: BIP-MAP Layer 1. This layer is constructed based on BPMN. The BI products generated at BIP-MAP Layer 2 are summarized onto this layer so that users are able to gain an overview of all the BI products that are available in the entire business process.

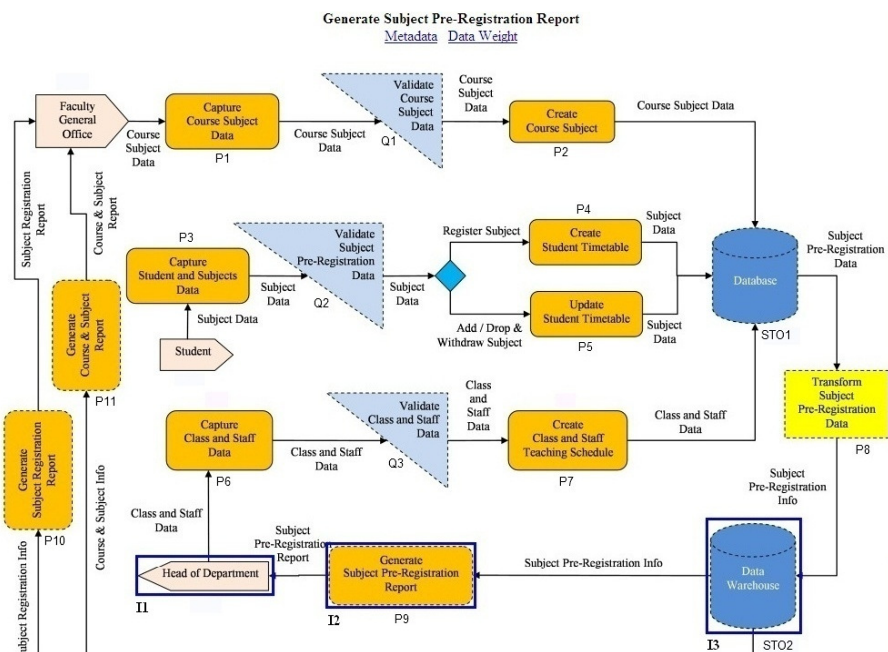


Figure 4.5: BIP-MAP Layer 2. This layer is constructed based on IP-MAP. Process B1 in Figure 4.4 is mapped to components I1, I2 and I3 where the student pre-registration report is generated by process P9 from data warehouse STO2 for the HODs.

Figure 4.4 shows the first layer of BIP-MAP for the subject pre-registration of FICT in UTAR. The first layer of BIP-MAP is constructed based on BPMN and it is used to model the business processes of an organization. In [3], BPMN breaks down a business process into a set of activities so that the management users are able to acquire a deeper understanding of the entire business process. In this thesis, the term 'activity' is substituted with the term 'process' at the business process layer in order to be standardized with the terminologies with the terminologies used for the information process layer in Section 4.4.2.

Integrating BPMN into a BI system has the following advantages. First, it enables the users to visualize a proper workflow of the processes by identifying its appropriate order and dependencies when a business process is segregated into multiple processes. Second, it enables the users to easily recognize the interaction between different participants and processes when they are grouped into separate sections with the swimlane object. It is important to clearly define all the processes and participants that are involved in a business process because an enterprise can be analyzed and integrated if its business processes are being modeled correctly [23]. When BPMN is integrated into a BI system, the BI products that are generated at the information process layer (BIP-MAP Layer 2) can be summarized onto the business process layer (BIP-MAP Layer 1) as an icon attached to a process. This enables users to gain an overview of 'what' BI products are available in the entire business process. Locating the BI products at the appropriate parts of BIP-MAP Layer 1 allows users to identify 'where' a BI product is produced and ‘when’ it is generated. Each BI product provides actionable information that serves a different purpose in decision making. As the users access the BI products, they are able to understand ‘why’ a BI product is required when the processes that are related to the BI product are represented in an orderly manner at BIP-MAP Layer 1. Thus, the integration of BPMN into a BI system helps users to update the BI products accordingly when there is a change in the business process so that the actionable information provided by the BI products is still relevant for the business operations and remains valid to be used for decision making.

BIP-MAP Layer 1 will be the main interface used by management users to navigate the proposed system where the information processes, data and BI products are organized based on business processes. Referring to Figure 4.4, when the management users click on any process with a dotted line boundary at BIP-MAP Layer 1, the process will be highlighted and a navigation bar that consists of a few links will appear on top of the screen. The *Information Process* link allows them to access the information process that is mapped to the business process and the *Dashboard* link allows them to access the BI products in various forms of graph or chart. The *Retrieve Data* link allows users to conveniently retrieve all the detailed data that is related to a process when it is necessary.

Figure 4.6 shows an example of the BI Dashboard that compares the projected student number and actual student number for a subject pre-registration in one of the trimesters. It helps the users to estimate a more accurate projection for the subject pre-registration of students based on the figures in the previous trimester. For example, the projected number of students will be highlighted in red if its figure is 10% less than the actual student number. If the projected student number is 10% more than the actual student number, it will be highlighted in yellow. Using the appropriate highlighting in the BI products will provide useful alerts to the users. Apart from this, they can also drill down from the BI Dashboard to view the data in a more detailed manner by clicking the individual graph bar of the actual student number. One of the examples of drilled down data is shown in Figure 4.7.

Figure 4.8 shows an example of the data that can be retrieved for a process when the users click on the *Retrieve Data* link at the navigation bar on top of BIP-MAP Layer 1. Allowing the users to retrieve the detailed data of a process will enable them to discover more data that can be utilized to generate useful BI products.

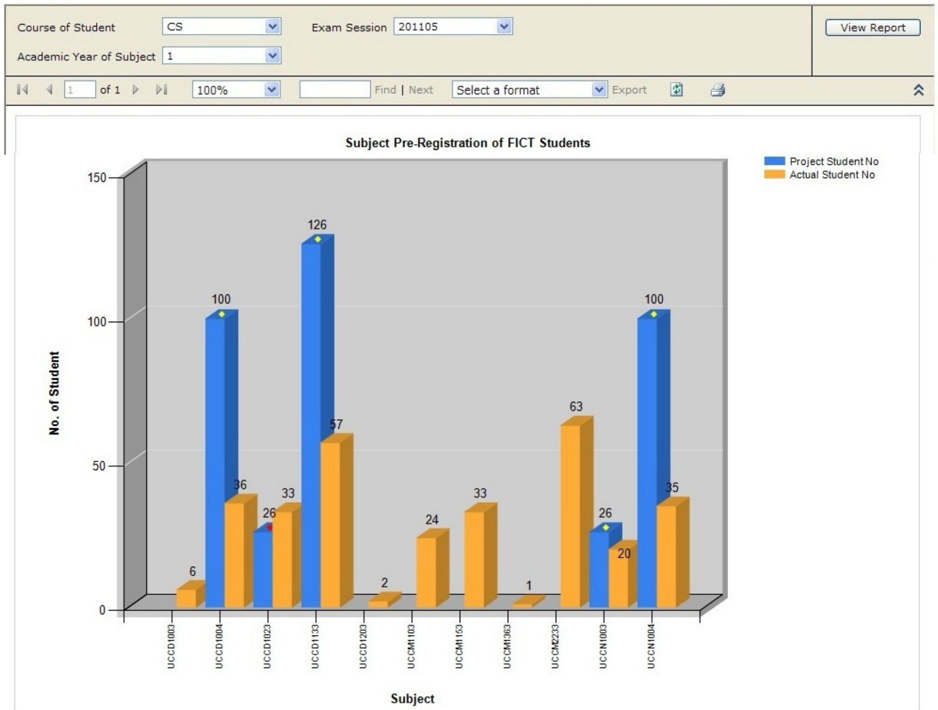


Figure 4.6: BI Dashboard. The BI products are presented to the users in various forms of dashboards where the users can drill down at different parts to retrieve more detailed information as shown in Figure 4.7.

UCCN1003 DATA COMMUNICATIONS AND NETWORKS

Student ID	Session Join	Structure Code
1M0AEBB	201001	UCCS100101
1M0DAPM	201005	UCCS100501
1M0PPDD	201005	UCCS100501
0V0BVVE	201101	UCCS110101
0V0EAVP	201101	UCCS110101
0V0PBZZ	201101	UCCS110101
0V0PDBB	201101	UCCS110101
0V0PDWE	201101	UCCS110101
0V0PVPE	201101	UCCS110101
1A0ABWB	201101	UCCS110101
1A0ABZP	201101	UCCS110101
1A0MZAV	201101	UCCS110101
1A0MZAW	201101	UCCS110101
1M0AZPB	201101	UCCS110101
0V0TEDW	201101	UCCS110101
0V0TVEZ	201101	UCCS110101
0V0ZEBE	201101	UCCS110101
0V0ZWVP	201101	UCCS110101
0W0DEBT	201101	UCCS110101
1A0AABD	201101	UCCS110101

Figure 4.7: Drilled Down Data of BI Product. Users are able to view detailed data of the BI product by clicking on different portions of the graph or chart.

Data Name	Data Weight
ActualStudentNo	0.3247
CourseCode	0.2225
ProjectStudentNo	0.2072
ExamSession	0.1058
AcademicYear	0.086
Trimester	0.0537
UnitCode	0

ExamSession	CourseCode	AcademicYear	Trimester	UnitCode	ActualStudentNo	ProjectStudentNo
201105	CN	1	1	UCCD1003	2	
201105	CN	1	1	UCCD1004	11	40
201105	CN	1	1	UCCD1133	12	40
201105	CN	1	1	UCCM1213	1	
201105	CN	1	1	UCCM2233	12	
201105	CN	1	1	UCCN1004	12	40
201105	CN	1	2	UCCD1023	13	13
201105	CN	1	2	UCCD1133	11	13
201105	CN	1	2	UCCM1213	11	
201105	CN	1	2	UCCN1003	11	13
201105	CN	1	3	UCCD1023	2	
201105	CN	2	1	UCCD2013	7	22
201105	CN	2	1	UCCD2203	4	22
201105	CN	2	1	UCCD3003	6	22
201105	CN	2	1	UCCD3113	21	22
201105	CN	2	1	UCCD3123	6	22
201105	CN	2	1	UCCD3213	19	22
201105	CN	2	1	UCCN2003	14	22
201105	CN	2	1	UCCN2023	17	22
201105	CN	2	1	UCCN2033	44	22
201105	CN	2	1	UCCN2103	18	22
201105	CN	2	1	UCCN2113	38	22
201105	CN	2	1	UCCN3213	19	22
201105	CN	2	2	UCCD2203	1	14
201105	CN	2	2	UCCD3003	1	14

1 2 3 4 5 6 7 8 9 10 ...

Figure 4.8: Retrieve Data Module. Users are allowed to conveniently retrieve all the detailed data that is related to a business process when it is necessary.

4.4.2 BIP-MAP Layer 2: Information Process Modeling for BI Product

The second layer of BIP-MAP is used to model the 'how' aspect of the BI system. Figure 4.5 shows the second layer of BIP-MAP for the subject pre-registration of FICT in UTAR. It is constructed based on IP-MAP [7] and being used to model the underlying information process that is utilized to generate the BI products within a business process. When the management users click on any process with a dotted line boundary at the first layer of BIP-MAP, the second layer of BIP-MAP will be invoked and the relevant information process for that particular business process will be highlighted.

By referring to the second layer of BIP-MAP, users are able to discover the mapping between a business process and its information process. Identifying the appropriate mapping of these two types of processes will help users to visualize the entire information manufacturing chain. When the information manufacturing chain of the BI products are properly modeled with the detailed data processing steps as described in Figure 4.5, users can easily identify how the data is being captured (P1, P3, P6), validated (Q1, Q2, Q3), processed (P2, P4, P5, P7), stored (STO1, STO2), transformed (P8) and generated (P9, P10, P11) throughout the organization. Apart from this, users are able to understand the system architecture of their organization. For example, they can identify the database and data warehouse that are utilized for data storage and recognize the data transformation steps for transferring the data from the database to the data warehouse. Having a proper understanding of the data processing steps will enable users to utilize the data for generating BI products that are more applicable to support their decision making. For instance, a manager who has a complete picture of the available data in the organization will be able to extract the appropriate data to be used for a specific case of decision making.

In the original IP-MAP [7], a standardized set of metadata is implemented at each building block for users to identify the department or role that is responsible in managing the data, location of the data, rules or procedures related to the data, data elements that compose the entire data set, and the base system (paper-based or electronic) of the data. But the metadata is not specifically categorized into different groups based on the functionality of each building block. If the data processing steps are not described in a specific manner with the relevant metadata, users will not be able to understand the details of the entire information manufacturing chain and this will cause the failure to use the right data for decision making. Therefore, in BIP-MAP, metadata is categorized into a few groups such as *Metadata for Data Validation*, *Metadata for Data Transformation*, *Metadata for Data Generation* and *Metadata for Database and Data Warehouse* to describe the different phases of a BI environment.

Referring to Figure 4.5, when the management users click on any component with a dotted line boundary at BIP-MAP Layer 2, a navigation bar that consists of two links will appear on top of the screen. The *Metadata* link allows users to access some information that describes the selected component. By referring to the metadata, both business users and technical users are able to gain a deeper understanding for the building blocks in the information manufacturing chain. In our approach, we differentiate different types of metadata for different information processes. The information processes are categorized based on the different stages of data processing, for example, data validation, data transformation, data generation and data warehousing.

Figure 4.9 shows an example of the *Metadata for Data Validation*. This metadata defines how data is being validated when it is being captured into the systems. As users know in detailed about how the data should be validated, it helps to capture data that is more useful into the systems. For example, referring to Figure 4.9, when performing data validation, users may need to find out what is the meaning of the various fields (Description), what is the valid range of inputs (Condition), how critical the data is (Compulsory), and the choices that are available for the data (Option). If any problem happens towards the validation of a set of data, users are able to identify the person (Data Steward) that is responsible in managing the data and its business rules. The solution for a business issue can be provided in a faster manner if the relevant person-in-charge for a set of data is identified easily. For instance, to validate the units to be offered for subject pre-registration, the field 'Unit Code' for a particular subject may be limited to a certain batch of students (Condition) and this field is critical (Compulsory) and the range of inputs for the 'Unit Code' field (Option) is the list of all subjects offered by the faculty under the administration of a faculty officer (Data Steward).

Metadata for Data Validation (Class and Staff Data)

Data	Description	Condition	Compulsory	Option	Data Steward
MID	- An auto-generated number used to uniquely identify a section which consists of one or more classes.	-		-	-
ClassID	- An auto-generated number used to uniquely identify a class.	-		-	-
Session	- The year and month of academic studies.	- Select one option only.		- List of Exam Sessions	-
Subject	- A subject to be taught in the class.			- List of Subjects Offered	Mr. Chan
Staff	- A lecturer or tutor who is teaching the subject.	- Select lecturers or tutors from the relevant fields of expertise.		- List of Lecturers and Tutors	Mr. Lee
Class Type	- The type of class being conducted for students.	- Select one option only.		- L : Lecture - T : Tutorial - P : Practical	-
Capacity	- The amount of seats available in a class.	- Numeric only.	Yes	-	-
Class Group	- The group number of a class.	- Numeric only.		-	-
Day of Week	- The day in which a class is being conducted.	- Select one or more options.		- 1 : Mon - 2 : Tue - 3 : Wed - 4 : Thu - 5 : Fri - 6 : Sat	-
Start Time	- The starting time of a class.	- Time format input only.		- A list of time from 8am to 7pm	-
End Time	- The ending time of a class.	- Time format input only.		- A list of time from 9am to 8pm	-
Venue	- The place in which a class is being conducted.	- Select one or more options.		- List of Venues	Mr. Lau

Figure 4.9: Metadata for Data Validation. The main objective of this metadata is to define the conditions that need to be fulfilled by the data and indicate whether the data is compulsory to be entered by the users.

Figure 4.10 shows an example of the *Metadata for Data Transformation*. This metadata serves as a useful reference for technical people like the database administrators and software developers since it provides the details for transformation of data from a database into a data warehouse. Understanding the data transformation process allows users to identify a better implementation of the Extract, Transform and Load (ETL) Process for data migration. For example, referring to Figure 4.10, users are able to identify the source (Database) and destination (Data Warehouse) of data, and understand the rules and procedures (ETL Process) that are involved in extracting, transforming and loading the data from a database into a data warehouse. For instance, to transform the subject pre-registration data from the database into the data warehouse, the database tables and type of storage for data (Database and Data Warehouse) are provided together with the procedures of counting the students (ETL Process) that have registered for each subject according to their exam session, course, academic year and trimester.

Metadata for Data Transformation

Database

SubjectPreRegistration Table	
exm_sess	int
stu_id	varchar(15)
unit_code	varchar(10)

CourseStructure Table	
crs_str	varchar(10)
unit_code	varchar(10)
unit_desc	varchar(255)
yr	int
tri	int

Student Table	
stu_id	varchar(15)
str_code	varchar(10)
crs_code	varchar(5)
sess_join	int

StudentProjectNo Table	
unit_code	varchar (10)
exm_sess	int
crs_code	varchar(5)
yr	int
tri	int
project_student_no	int

ETL Process

1. Count the number of students that have registered for each subject by exam session, course, academic year and trimester.

2. Data Extraction Rule:
- stu_status = 'EN'

3. Data Storage Location:
- unit_code => UnitCode
- exm_sess => ExamSession
- crs_code => Course
- yr => AcademicYear
- tri => Trimester
- COUNT(stu_id) => ActualStudentNo
- project_student_no => ProjectStudentNo

Data Warehouse

FactStudentFigure Table	
UnitCode	varchar (10)
ExamSession	int
Course	varchar(5)
AcademicYear	int
Trimester	int
ActualStudentNo	int
ProjectStudentNo	int

Figure 4.10: Metadata for Data Transformation. The main objective of this metadata is to provide a detailed set of rules and procedures that are involved in the data transformation process so that users can improve the implementation of ETL process for data migration.

Figure 4.11 shows an example of the *Metadata for Data Generation*. This metadata describes the steps and criteria that are used to generate a BI product from the systems for reporting. For example, referring to Figure 4.11, management users are able to identify the default rule and optional rule (Report Generation Process) that are involved in generating the data of a BI product, and recognize how the data is stored (Data Warehouse) and presented (Report Structure) in the BI Dashboard. This enables them to identify additional data that may be more applicable to their decision making. For instance, to generate the subject pre-registration data from the data warehouse, the default rule is to calculate the actual number and projected number of the Year 1 Computer Science students for all the subjects in the previous trimester. Optionally, users are also allowed to select the subject pre-registration data of the students for other courses, trimesters and academic years.

Metadata for Data Generation (Subject Pre-Registration)

Data Warehouse			Report Generation Process	Report Structure	
FactStudentFigure			Report Objective: 1. To calculate the actual number of students who have registered for each subject and its projected number so that class capacity can be allocated accordingly. Default Data Selection Rule: 1. SUBSTRING(UnitCode,2,1) = 'C' 2. Course = 'CS' 3. ExamSession = 'Previous Trimester' 4. AcademicYear = '1' Optional Data Selection Rule: 1. Course = 'CT, CN, IA, IB' 2. ExamSession = 'Other Trimesters' 3. Academic Year = '2, 3'	*Chart Title*	Subject Pre-Registration of FICT Students
UnitCode	varchar (10)	Refer to Subject List		*Chart Type*	Bar Graph
ExamSession	int	-		*Dimension*	2D
AcademicYear		-		*X-Axis*	Unit Code
Trimester		-		*Y-Axis*	SUM(ProjectStudentNo) SUM(ActualStudentNo)
ActualStudentNo		-			
ProjectStudentNo		-			

Figure 4.11: Metadata for Data Generation. The main objective of this metadata is to enable users to understand the steps and criteria used for generating a BI product so that additional data that is more applicable to their decision making can be identified easily.

Figure 4.12 shows an example of the *Metadata for Database and Data Warehouse*. This metadata serves as a useful reference for users to improve the corporate data governance when a detailed description of each data attribute is provided to them. With the use of metadata in data governance, companies are able to implement accountabilities to manage the quality of data at a corporate-wide level [74]. For example, referring to Figure 4.12, users are able to identify the person (Data Custodian) that is responsible in managing the data and its technical rules. The solution for a technical issue can be provided in a faster manner if the relevant person-in-charge for a set of data is identified easily. Apart from this, users are able to find out what is the meaning of each field (Description), what are the various input methods of data (Input Type), where does the data come from (Data Source), how it is being stored (Data Type) and what it is representing (Data Code). It is essential for users to understand the meaning of data so that the data will not be misinterpreted for decision making.

CourseStructure

Data	Description	Input Type	Data Source	Data Type	Data Code	Data Custodian
crs_str	- A code that represents the course structure.	Form Input	Structure Code	varchar (10)	-	Mr. Chong
unit_code	- A code that represents the subject.		Unit Code	varchar (10)	Refer to Subject List	
unit_desc	- A name that describes the subject.		Unit Description	varchar (255)	-	
yr	- A year of academic studies for the subject.		Year	int	-	-
tri	- A term of academic studies for the subject.		Trimester	int	-	-

FactStudentFigure

Data	Description	Input Type	Data Source	Data Type	Data Code	Data Custodian
UnitCode	- A code that represents the subject.	Extraction	- Data File * UnitCode	varchar (10)	Refer to Subject List	Mr. Chong
ExamSession	- The year and month of academic studies.		- Data File * ExamSession	int	-	-
AcademicYear	- A year of academic studies.		- Data File * AcademicYear		-	-
Trimester	- A term of academic studies		- Data File * Trimester		-	-
ActualStudentNo	- The actual number of students that registered for the subject.		- Data File * ActualStudentNo		-	-
ProjectStudentNo	- The projected number of students that registered for the subject.		- Data File * ProjectStudentNo		-	-

Figure 4.12: Metadata for Database and Data Warehouse. The main objective of this metadata is to provide a detailed description of each data attribute in the database and data warehouse so that users are able to implement data governance at a corporate-wide level.

4.5 Evaluation and Discussion

In order to confirm the usefulness of integrating the BPMN and IP-MAP models, two surveys have been conducted to gather some feedback from the management users. Section 4.5.1 discusses the feedback of survey conducted at Universiti Tunku Abdul Rahman and Section 4.5.2 discusses the feedback of survey conducted at an online job recruitment firm.

4.5.1 Case Study I

BIP-MAP is compared against the original BPMN and IP-MAP models for the subject pre-registration of FICT in UTAR. A total of ten members from the faculty management committee have evaluated the framework. The evaluators are briefed about the functionalities for each of the three models. Then, they answered a series of questions to evaluate the performance of each model. The survey questionnaire has been devised carefully to verify the usefulness of integrating BPMN and IP-MAP and to show how the different features in the two models complement each other. It is attached in Appendix A.

Table 4.6 provides a brief summary of the features that are evaluated for the three models with an average rating from all the ten evaluators. The features of the three models are evaluated one after another using the five-point likert scale. As the evaluators rated the usefulness of each feature in BPMN, they checked whether the same feature is available in IP-MAP. Similarly, as they evaluated the features in IP-MAP, they also checked whether the same features are available in BPMN. In other words, cross-checking is performed in order to determine whether the two models complement each other.

Table 4.6: Features Evaluated in BPMN, IP-MAP and BIP-MAP

BPMN		
Feature	**Usefulness**	**Availability in IP-MAP**
1. Process Segregation Breaks down a business process into a set of sub-processes.	✓ (4.30)	✓ (4.00)
2. Workflow Visualization Helps users to visualize the workflow of processes and identify the appropriate order and dependencies among the processes.	✓ (4.50)	○ (3.67)
3. Users Interaction Allows users to easily identify the interaction between all participants that are involved in a business process.	○ (3.80)	✗ (2.56)

IP-MAP		
Feature	**Usefulness**	**Availability in BPMN**
1. Information Product Enable users to visualize all the information products in an information process.	✓ (4.10)	✗ (1.67)
2. System Architecture Helps users to understand how data is processed and stored in the database.	○ (3.70)	✗ (2.78)
3. Data Processing Allows users to visualize the data processing steps to generate the information reports.	○ (3.8)	✗ (1.56)

BIP-MAP (Additional Features in BIP-MAP)	
Feature	**Usefulness**
1. Process Mapping Mapping the business processes (BPMN) with information processes (IP-MAP).	○ (3.80)
2. BI Dashboard Bringing the information product (BI Dashboard) from the BIP-MAP Layer 2 (IP-MAP) into Layer 1 (BPMN).	✓ (4.00)
3. Data Retrieval Retrieving and viewing detailed data related to a business process and an information product at the 1st layer of BIP-Map.	✓ (4.10)

<u>Symbol</u>	<u>Range of Average Rating</u>
✓	4.00 – 5.00 (Agree / Strongly Agree)
○	3.00 – 3.99 (Neutral)
✗	1.00 – 2.99 (Strongly Disagree / Disagree)

The three features that are evaluated for BPMN include process segregation, workflow visualization and users interaction. The process segregation feature is rated with a value of 4.3 for BPMN and 4 for IP-MAP. This showed that both models have been properly designed to break down a process into a set of sub-processes so that users are able to easily identify the detailed steps of a process. For the workflow visualization feature, BPMN is rated with a value of 4.5 but IP-MAP is only rated with a value of 3.67. This is because BPMN enable users to identify the appropriate order and dependencies among the processes. On the other hand, IP-MAP does not model the processes in a chronological order and the dependencies among the processes are also not shown. Besides, users only gave a moderate rating of 3.8 for the user interaction feature of BPMN. This is probably the interviewed users are already familiar with the workings of the subject pre-registration process and do not think that such documentation is necessary. The value is still higher compared to IP-MAP with a value of 2.56 where the owners of the various information processes are not specified and the interaction between different participants is not documented or clearly defined especially when it involves different departments.

For IP-MAP, the three features that have been evaluated are related to the information product, system architecture and data processing. IP-MAP is rated with a value of 4.1 for the information products in an information process whereas BPMN is only rated with a very low value of 1.67. This is because BPMN focuses on modeling the business process layer instead of the information process layer in which the information products are generated. In addition, the users are neutral about being able to comprehend the system architecture and the data processing steps to generate the information products (3.7 and 3.8 respectively). One of the possible reasons is that they are keen to understand each business process in terms of all kinds of information processing steps, and not only confined to those related to the generation of information products. These two features are not available in BPMN (2.78 and 1.56, respectively).

From the evaluation results of BPMN and IP-MAP, it is apparent that most of the features available in both of the models are not overlapping with one another. The functions available in one model are not able to be performed completely by the other. For example, IP-MAP allows users to visualize the detailed data processing steps that are utilized to generate the information products in an organization but BPMN does not model the information processes that are involved in manipulating the data. On the other hand, BPMN allows users to identify the appropriate order and dependencies among the business processes but IP-MAP does not model the business processes that are closely related to the information products of an organization. Therefore, the features of BPMN and IP-MAP indeed complement each other and it is necessary to integrate them in order to construct a robust framework for BI.

As BPMN and IP-MAP are integrated to construct BIP-MAP, the three main features that are evaluated for the integration of the two models are process mapping, BI Dashboard and data retrieval. The feature of process mapping where the business process and information process are integrated with one another is rated with a moderate value of 3.8 because one of the users felt that BIP-MAP should provide the details of some documents to the users rather than just mapping the processes together. The features of BI Dashboard and data retrieval module are rated with a high value of 4 and 4.1 respectively. This is because the BI Dashboard enables users to view the important information in a visual manner using various kinds of graphs or charts while the data retrieval module allows them to conveniently retrieve very specific data that is used to generate the information products.

Furthermore, the evaluators provided a score to rate the overall performance of the three models and ranked them accordingly. The results of evaluation for the overall performance of the three models are shown in Figure 4.13. The evaluated features are as follows:

- **Business Process Management**: To break down the processes into sub tasks by having them to be represented by different modules so that all the processes can be managed easily.
- **User Interaction**: To identify the roles of participants and their interactions with one another for different tasks.
- **Business Intelligence Reporting**: To identify the intermediate and final deliverables at different phases of a process.
- **Information Process Management**: To model the processes that describe how data is being captured, validated, processed, stored and utilized to generate the information reports.

From the comparison of the three models, it is found that BPMN and IP-MAP provides a better functionality for different features. While BPMN performs well for *Business Process Management* and *User Interaction*, IP-MAP performs well only for *Information Process Management.* This shows that both models can only present a partial view of the business operations. BPMN can represent business process more clearly by showing the flow of activities and interaction of users with one another. On the other hand, IP-MAP provides a better feature for managing information process because the flow of data is modeled where how the data is manipulated in different processes is clearly shown. Both models do not perform well on *Business Intelligence Reporting* since the BI products are not graphically presented to the users for business analysis and decision making. When the salient modeling techniques of BPMN and IP-MAP are integrated into BIP-MAP, BIP-MAP outperforms both models with the highest score in ***all*** features. This shows that BIP-MAP is able to give a more comprehensive view of the business operations when users are capable of visualizing the entire business management workflow and information manufacturing chain.

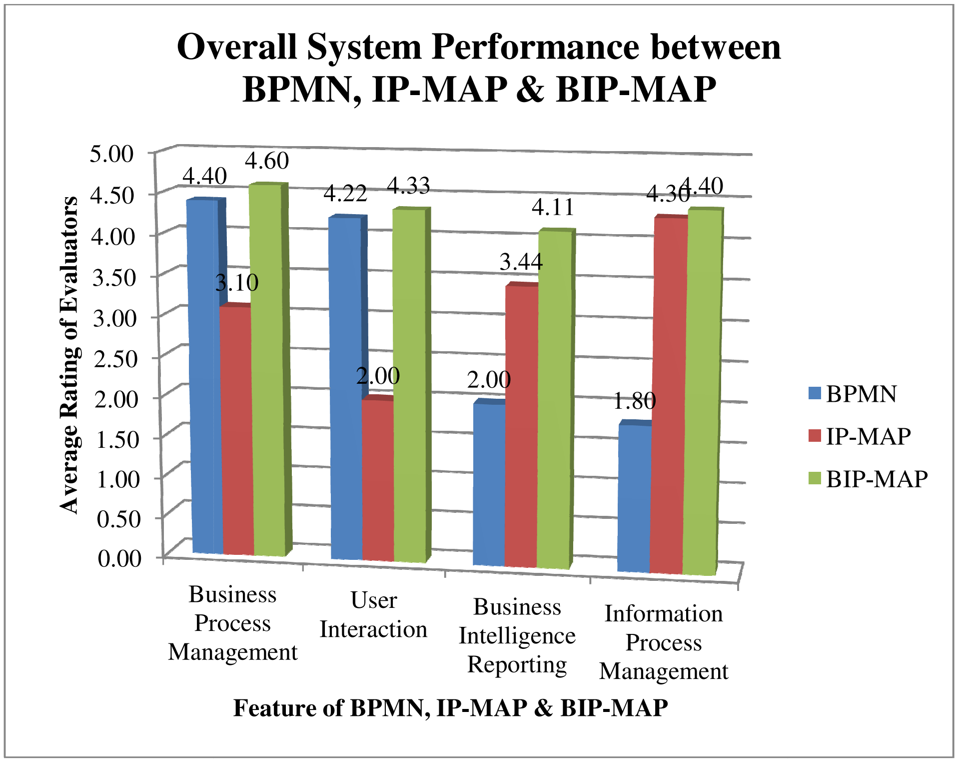


Figure 4.13: Overall System Performance between BPMN, IP-MAP and BIP-MAP (proposed framework). The features in the three models are rated by the evaluators in a comparative manner.

The evaluators provided their overall feedback for BIP-MAP by answering some general questions which have been attached as a reference in Appendix A. Majority of them agreed that the integration of BPMN and IP-MAP into BIP-MAP is useful in assisting users to perform the decision making process in an organization because of the following reasons:

- BIP-MAP offers a more comprehensive view of the business process and information process as users are allowed to click at a particular process to retrieve more details.
- The business process layer (BIP-MAP Layer 1) has provided them a clear picture about the business processes that are involved in producing the BI products used for decision making.

- The information process layer (BIP-MAP Layer 2) has provided them a clear picture about how the data is being captured, validated, processed, stored, transformed and generated throughout the organization.

The evaluators also gave the following feedback to further enhance BIP-MAP:

- **Suggesting Possible Solutions to a Problem**
  - ✓ One of the evaluators indicated that BIP-MAP is only useful for identifying the possible areas of problem for a BI product but it is not completely helpful in assisting users to make decisions because the options of solution for a particular problem are not provided to the users. The appropriate solution for a particular problem always needs to be formulated from a case to case basis. BIP-MAP can be enhanced by including some possible options of solution towards any problem which may happen for the BI products so that users are able to choose an option of solution to rectify a particular problem.
- **Incorporating Retrieve Data Module with BI Dashboard**
  - ✓ Another evaluator felt that the Retrieve Data Module (Figure 4.8) is redundant and it should be incorporated into the BI Dashboard in order to give a clearer and more impressive view of the data. This module is not redundant since not all data is included in the BI Dashboard because the purpose of BI Dashboard is to deliver important information to users in a summarized manner. Therefore, this module is necessary to enable users to retrieve the detailed data of a particular process when there is a need for them to do so.

- **Providing Additional Information about Document**
  - ✓ It is suggested by one of the evaluators that the exact format of any document should be provided to the users rather than just list out the detailed processes that manipulate the data. BIP-MAP can be enhanced by including some additional information that describes the details of any document that is related to a business process so that users are able to utilize the document appropriately.
- **Implementing a "What-If" Data Analysis Feature**
  - ✓ Last but not least, one of the evaluators proposed to implement a "What-If" analysis feature for the BI Dashboard so that users are able to retrieve the data based on different criteria in a more flexible manner. The current BIP-MAP allows users to search the data with different criteria. Its search function can be improved by implementing the "What-If" data analysis feature so that users are able to view the data in different perspectives.

4.5.2 Case Study II

An additional survey is performed for BIP-MAP at an online job recruitment firm that allows the job seekers to search for various kinds of jobs offered by the companies when both parties registered themselves as members of the organization. After registering themselves as members, job seekers can post their resume online and companies can post the details of the jobs offered by them online. BIP-MAP has been modeled for three different types of processes that relate to the job seekers, job advertisements and job competitors respectively. Figure 4.14 to Figure 4.21 shows the BI Dashboard and BIP-

MAP of the job seekers process which includes the business process layer, information process layer and metadata. The screenshots for the processes of job advertisements and job competitors are attached as references in Appendix C and Appendix D.

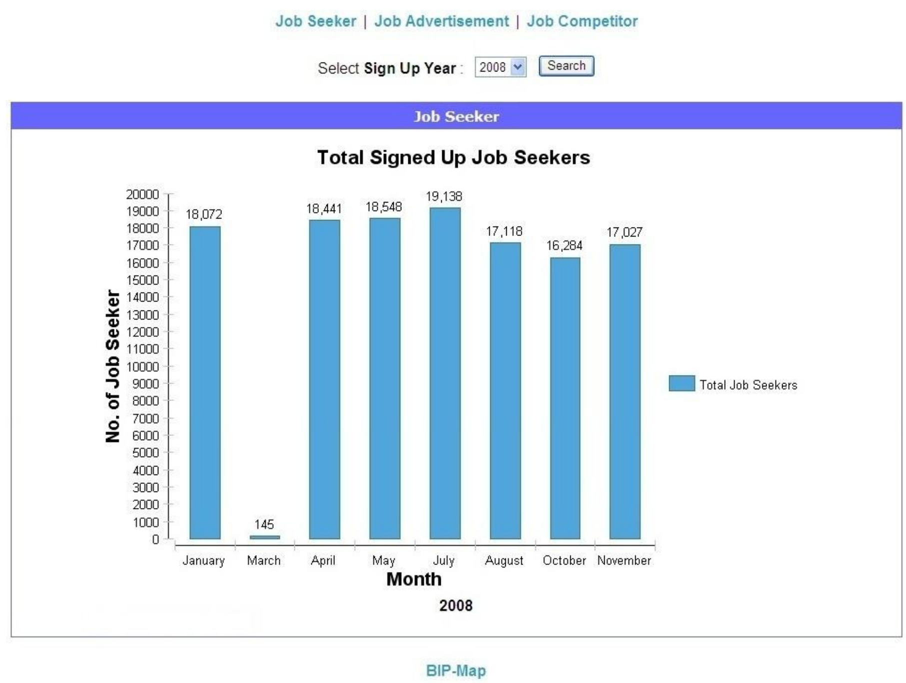


Figure 4.14: BI Dashboard (Job Seeker). It presents to the users the total number of job seekers that have signed up as members of the online job recruitment firm in each month of the year.

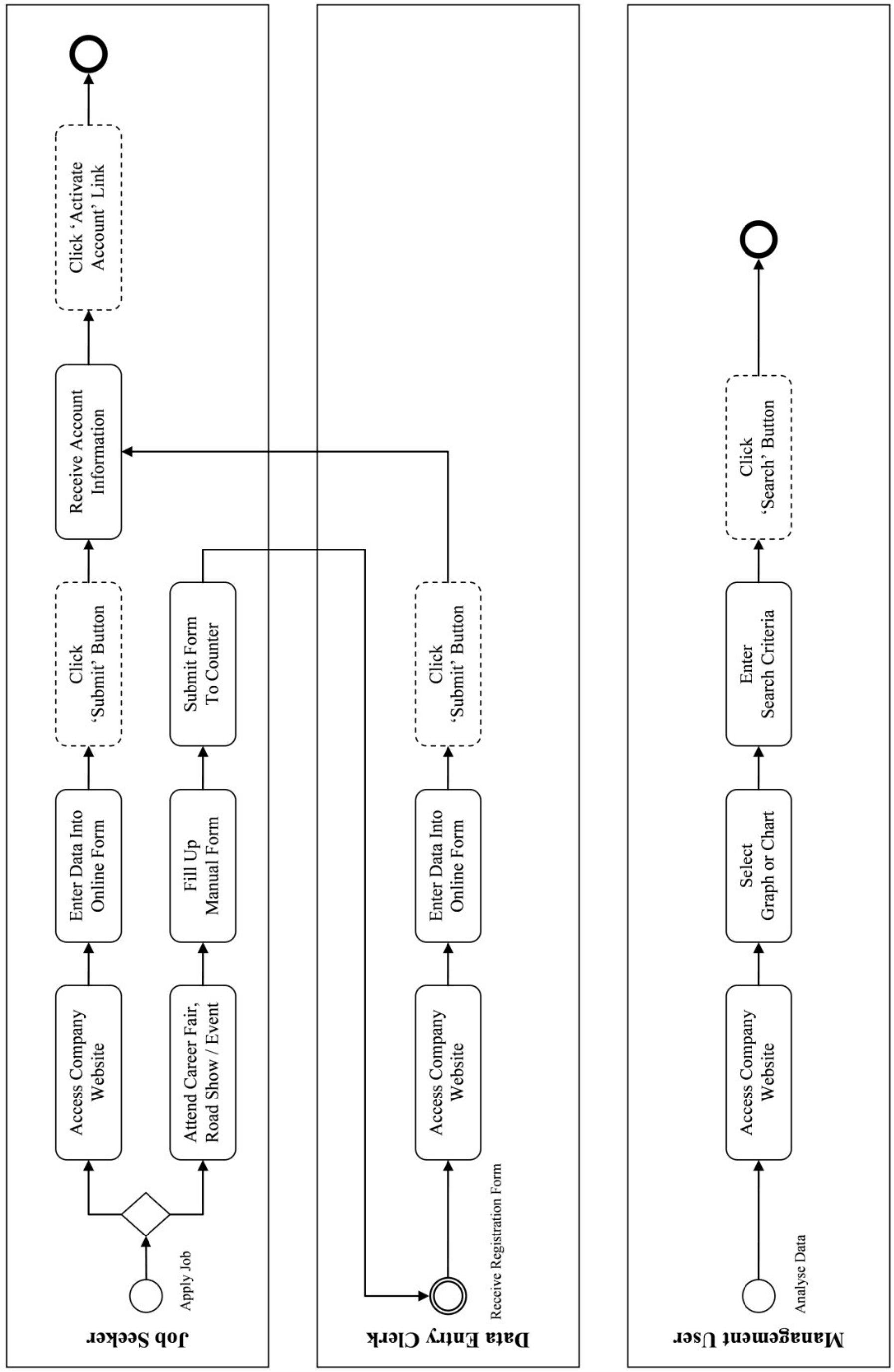


Figure 4.15: BIP-MAP Layer 1 (Job Seeker). This layer is constructed based on BPMN. It presents to the users the business processes for a job seeker to sign up as a member of the online job recruitment firm.

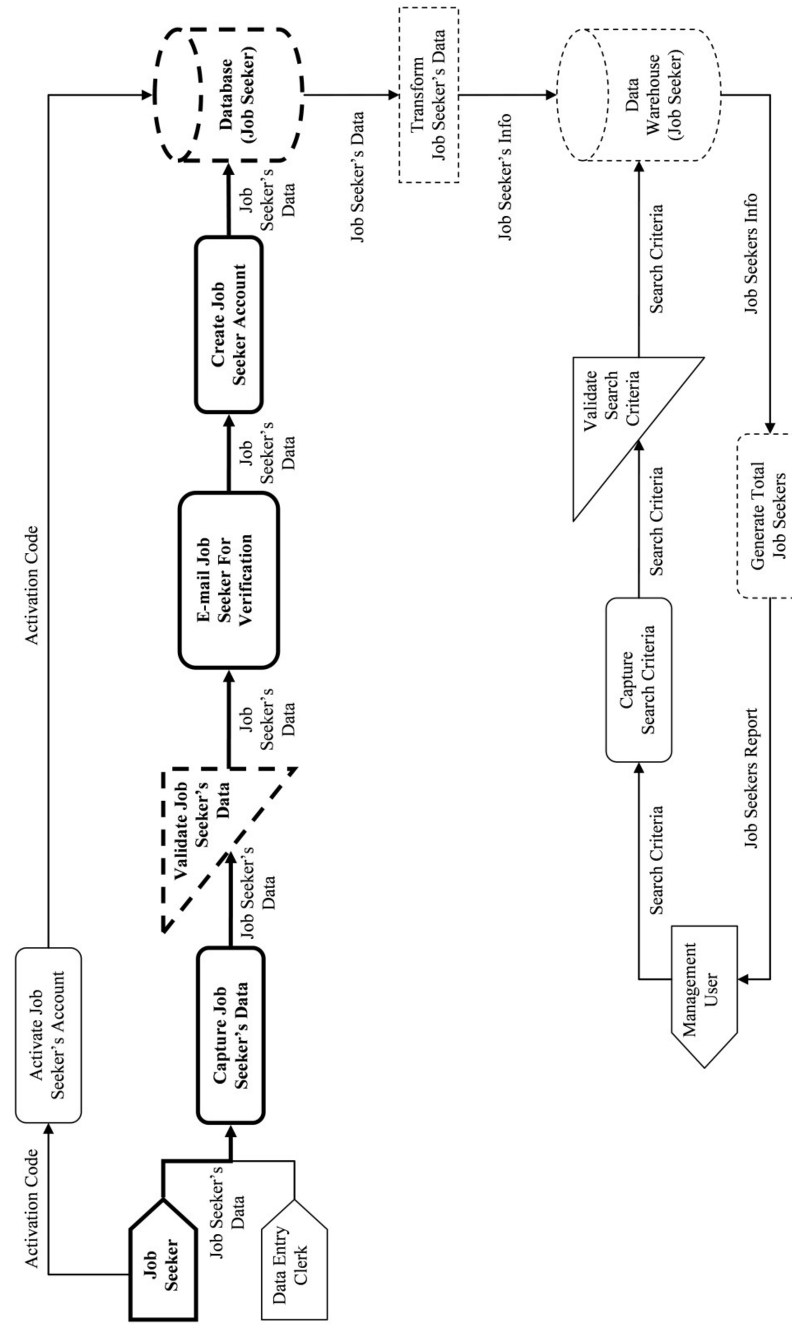


Figure 4.16: BIP-MAP Layer 2 (Job Seeker). This layer is constructed based on IP-MAP. It presents to the users the information processes for registering a job seeker as a member of the online job recruitment firm.

Metadata for Data Validation (Job Seeker)

Data	Description	Condition	Compulsory	Option	Data Steward
E-mail	– E-mail address is used as the account login ID.	– Any valid e-mail address.	Yes	-	Mr. Lee
Password	– A string of letters or numbers used for authentication with the e-mail.	– 6 to 60 characters. – Combination of letters and numbers with lowercase and uppercase (optional).		-	
Name	– First Name and Last Name of the user.	– Number is not allowed.		-	
Nationality	– The country to which the user belongs.	– Select one option only.		– A list of countries.	Mr. Chan
Country of Residence	– The country where the user is staying.				
Working Location – Where do you want to work	– The place where the user prefers to work.			– A list of working locations.	
Position Level – Level of position you want to work in	– The position level where the user prefers to work.	– Select one or more options.		– Senior Manager – Manager – Senior Executive – Junior Executive – Fresh / Entry Level – Non-Executive – Student Seeking Internship	Mr. Lau
Specialization – What do you want to specialize in (Maximum 5)	– The field of work in which the user has been trained with extra knowledge.			– A list of job specialization fields.	
Source – I learned about JobStreet.com from	– The place where the user learned about the company.	– Select one option only.		– A list of advertisement methods.	Mr. Lee

Figure 4.17: Metadata for Data Validation (Job Seeker). The main objective of this metadata is to define the conditions that need to be fulfilled by the data and indicate whether the data is compulsory to be entered by the users.

Metadata for Data Transformation (Job Seeker)

Database		ETL Process	Data Warehouse	
hst_central_candidate Table		1. Set the current system date as the Reporting Date (ReportDate). • Current System Date → ***ReportDate*** 2. Set the months and years selected by users as the Sign Up Date (SignUpDate) to identify the total number of job seekers by verifying their Submission Date (submit_date). • Selected Months and Years → ***SignUpDate*** 3. Count the job seekers that signed up within the selected Sign Up Date for each Country (country_code) and E-mail Status (email_status_code). • Data Extraction Rule : ○ ***account_status_code*** = 1 • Data Storage Location : ○ ***country_code*** → ***CountryCode*** ○ ***email_status_code*** → ***EmailStatusCode*** ○ Total Job Seekers → ***TotalCandidate***	**CandidateSignUp Table**	
global_candidate_id	int		ReportDate	datetime
email	varchar(50)		CountryCode	int
password	varchar(60)		SignUpDate	datetime
name	varchar(50)		EmailStatusCode	int
nationality	int		TotalCandidate	
country_code				
working_location				
position	varchar(50)			
specialization				
source	int			
account_status_code				
email_status_code				
submit_date	datetime			

Figure 4.18: Metadata for Data Transformation (Job Seeker). The main objective of this metadata is to provide a detailed set of rules and procedures that are involved in the data transformation process so that users can improve the implementation of ETL process for data migration.

Metadata for Data Generation (Job Seeker)

Data Warehouse			Report Generation Process	Report Structure	
CandidateSignUp Table			Report Objective : 1. To calculate the number of users which have registered themselves as job seekers. Default Data Selection Rule : 1. Year(***SignUpDate***) = Sign Up Year 2. Job Seekers in Malaysia (***CountryCode*** = MY) 3. Job Seekers with Verified Account (***EmailStatusCode*** = 1, 3, 5) Optional Data Selection Rule : 1. Year(***SignUpDate***) = Sign Up Year 2. Job Seekers in Overseas Countries (***CountryCode*** != MY) 3. Job Seekers without Verified Account (***EmailStatusCode*** = 2, 4, 6)	Chart Title	Total Number of Job Seekers
ReportDate	datetime	-		Chart Type	Bar Graph
CountryCode	int	– Refer to Country List		Dimension	2D
SignUpDate	datetime	-		X-Axis	SignUpDate (Month)
EmailStatusCode	int	1, 3, 5 : Valid 2, 4, 6 : Invalid		Y-Axis	TotalCandidate
TotalCandidate	int	-			

Figure 4.19: Metadata for Data Generation (Job Seeker). The main objective of this metadata is to enable users to understand the steps and criteria used for generating a BI product so that additional data that is more applicable to their decision making can be identified easily.

Metadata for Database (Job Seeker)

hst_central_candidate						
Data	**Description**	**Input Type**	**Data Source**	**Data Type**	**Data Code**	**Data Custodian**
global_candidate_id	– An auto-generated number used to uniquely identify a job seeker.	System Input	-	int	-	-
email	– E-mail address is used as the account login ID.	Form Input	E-mail	varchar(50)	-	-
password	– A string of letters or numbers used for authentication with the e-mail.		Password	varchar(60)	-	-
name	– First Name and Last Name of the user.		Name	varchar(50)	-	-
nationality	– A country that the user belongs to.		Nationality	int	– Refer to Country List	Mr. Chong
country_code	– A country that the user is staying.		Country of Residence			
working_location	– A place that the user prefers to work.		Working Location		– Refer to Location List	
position	– A position level that the user prefers to work.		Position Level	varchar(50)	1 : Senior Manager 2 : Manager 3 : Senior Executive 4 : Junior Executive 5 : Fresh / Entry Level 6 : Non-Executive -1 : Student Internship	Mr. Kok
specialization	– A field of work that the user is trained with extra knowledge.		Specialization		– Refer to Specialization List	
source	– A place where the user learned about the company.		Source	int	– Refer to Source List	Mr. Lai
account_status_code	– Indicates whether the account of a user is active or not.	User Input	-		1 : Active 0 : Inactive	-
email_status_code	– Indicates whether the account of a user has been activated through e-mail.		-		1, 3, 5 : Valid 2, 4, 6 : Invalid	-
submit_date	– A date that the user signed up as job seeker online.	System Input	-	datetime	-	-

Figure 4.20: Metadata for Database (Job Seeker). The main objective of this metadata is to provide a detailed description of each data attribute in the database so that users are able to implement data governance at a corporate-wide level.

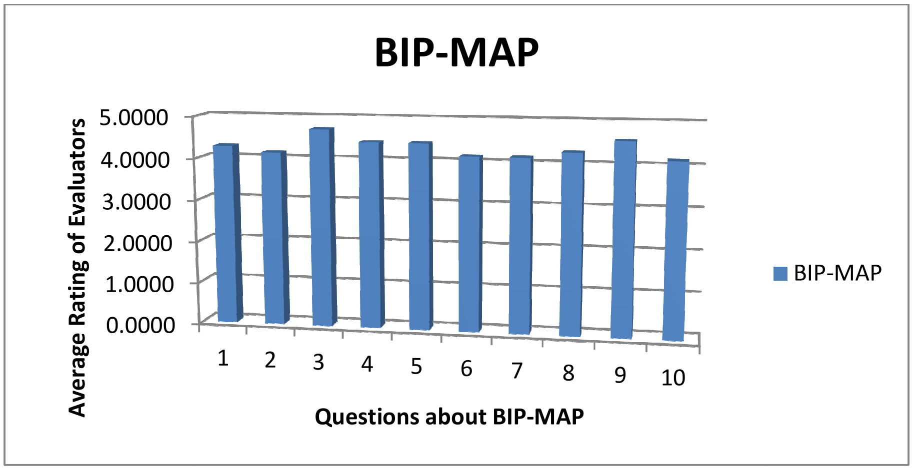
Metadata for Data Warehouse (Job Seeker)

CandidateSignUp						
Data	**Description**	**Input Type**	**Data Source**	**Data Type**	**Data Code**	**Data Custodian**
ReportDate	– A date that the data is loaded into the Data Warehouse.	Generation	– Auto-generated based on the current system date.	datetime	-	-
CountryCode	– A country that the users are staying.	Extraction	– hst_central_candidate * country_code	int	– Refer to Country List	Mr. Chong
SignUpDate	– A date that the users signed up as job seekers online.	Generation	– Auto-generated based on the months and years selected for the ETL process.	datetime	-	-
EmailStatusCode	– Indicates whether the account of users have been activated through e-mail.	Extraction	– hst_central_candidate * email_status_code	int	1, 3, 5 : Valid 2, 4, 6 : Invalid	-
TotalCandidate	– The total number of job seekers.	Calculation	– Calculated by summing up the job seekers that matched the CountryCode, SignUpDate and EmailStatusCode.		-	-

Figure 4.21: Metadata for Data Warehouse (Job Seeker). The main objective of this metadata is to provide a detailed description of each data attribute in the data warehouse so that users are able to implement data governance at a corporate-wide level.

Similarly, BIP-MAP is evaluated by seven members at the management level in the online job recruitment firm. The functionalities of BIP-MAP are presented to the evaluators and they assessed the framework by answering two set of general questions. The first set of questions allow them to provide their feedback about BIP-MAP using the five-point likert scale and the second set of questions allow them to provide some general comments for BIP-MAP. Both of the questionnaires are attached as a reference in Appendix B. Figure 4.22 shows the results of evaluation for BIP-MAP based on the first set of questions. It is evident that majority of the evaluators either agreed or strongly agreed on the usefulness of BIP-MAP because of the following reasons:

- BIP-MAP is useful for business decision making when it is being integrated with a BI Dashboard because it provides the management users a clear picture of the information manufacturing processes and also understanding of how the data is being tabulated into the various BI products of an organization.
- BIP-MAP will enable the management users to gain more insight to the displayed reports or charts in a BI Dashboard.
- BIP-MAP can be used to improve organizational business operations and help to determine the type of reports that are required.
- It is good to have a combination of the two layers in BIP-Map because the technical users are able to understand the business considerations behind certain data, while the business users can get an insight into the logic behind the information generated to them.



No.	Description of Question
1	BI Dashboard is useful in providing information to Managers and Business Analysts during the decision making process of an organisation.
2	Managers and Business Analysts may possess limited picture about the entire information manufacturing process underneath a particular piece of graph or chart produced in the BI Dashboard.
3	It is good to have a framework that can allow users to identify the business processes that are related to the information provided to them.
4	It is good to have a framework that can allow users to visualise the manufacturing processes of information in their organisations especially from a quality aspect.
5	It is good to have a framework that can allow users to understand the real meaning of data by referring to the metadata available in the framework.
6	Having such a framework would allow users to easily identify the critical phases that may create data quality problems.
7	Having such a framework would allow users to implement continuous improvement for the BI needs of their organisations.
8	The BIP-Map is useful since it provides the workflow of business processes so that users can understand the relationship of all participants and processes for a business transaction in order to implement Business Process Management.
9	The BIP-Map is useful since it provides the workflow of information processes so that users can identify how the data is being captured, processed, stored and utilised in order to produce better information products.
10	The BIP-Map is useful since it provides metadata so that users can understand the real meaning of the data before using it to make important decision for their organisations.

Figure 4.22: Evaluation of BIP-MAP. Majority of the evaluators either agreed or strongly agreed on the usefulness of BIP-MAP.

The evaluators also gave the following feedback to further enhance BIP-MAP:

- **Implementing BIP-MAP for Complex Systems**
  - ✓ A similar concern raised by a few evaluators is the challenge to implement BIP-MAP for complex systems, especially those that have ongoing changes because the creation and update for BIP-MAP will be problematic. Even though it may not be easy to model all the business and information processes, it is surely worth it to construct a BIP-MAP because it will serve as a helpful tool for improving the BI needs of an organization. In addition, the ongoing changes of business and information processes for complex systems can also be updated easily by referring to the BIP-MAP.
- **Implementing a "What-If" Data Analysis Feature**
  - ✓ One evaluator proposed to implement a feature for the BI Dashboard to automate and suggest the possible directions of data output based on different data input. This feature is similar to the "What-If" data analysis feature proposed by one of the evaluators in the first case study. The current search function of BIP-MAP can be improved by including the "What-If" data analysis feature.
- **Providing Metadata to the Relevant Users**
  - ✓ Two evaluators felt that metadata should be accessed by the business analysis or report generator team only as the information provided may be overloaded or taxing for the management users to understand. However, metadata is able to provide the detailed information that describes a BI product if there is a need for the management users to

further understand the business process and information process that are related to a BI product.

- **Constructing BIP-MAP with a Software**
  - ✓ One of the evaluators suggested that having a software to automatically plot the different layers based on user selections would be very useful especially if the software can be linked to the actual coding of the reports in order to increase efficiency and accuracy of maintaining the BIP-MAP. Indeed it is better to have a software that can allow users to construct the BIP-MAP in a flexible manner where users are able to build the business process and information process layers, manipulate the BI products and manage the metadata easily.
- **Implementing Control Management for Information Security**
  - ✓ Last but not least, one evaluator proposed to implement a control management feature in BIP-MAP for ensuring information security. BIP-MAP can be enhanced to implement information security by including the access control feature to deliver the right information to the right users.

4.6 Chapter Summary

This chapter has presented the functionalities of BIP-MAP that are integrated and enhanced from BPMN and IP-MAP. Two case studies have been performed to evaluate the usefulness of BIP-MAP, one in UTAR and the other one in an online job recruitment firm. Both of the case studies confirmed that it is necessary to implement a framework which integrates the business and information processes in order to provide actionable information to users.

CHAPTER 5

IMPLEMENTATION OF DATA QUALITY

5.1 Introduction to Data Quality

In a competitive business environment, one of the most important issues that need to be addressed by an organization is its data quality. This is because the usage of high quality information will enable the management users to make correct decisions in their daily business. As a result, most organizations have spent a great investment in technology to capture, store and process a huge amount of data and convert them into meaningful information and knowledge so that they can succeed over their competitors [75].

One effective way to deal with the issues of data quality is to implement the concept of Total Data Quality Management (TDQM) [62] where information is managed as products just like the products being manufactured in any industry. The ultimate aim of TDQM is to provide high quality information products that satisfy the needs of all users in an organization. In the cycle of TDQM, information quality is being defined, measured, analyzed and improved continuously to ensure high quality information products are delivered to all the information users.

Information Product Map (IP-MAP) [64] is a model that is developed to implement TDQM by providing some algorithms to compute the basic data quality dimensions like *completeness*, *accuracy*, and *timeliness* of data. The completeness [76] of a data element is the degree to which a value is available for that particular data element, and the completeness of data is determined by computing the availability of a value for each data element in the information product. However, an information product with some data elements that have missing values is still recognized as complete as long as the data elements available in the information product are adequate for the case of decision making [64]. The accuracy [77] of a data element is the degree of closeness for the value of that particular data element to the value of a corresponding one, and the accuracy of data is determined by computing the percentage of difference for a correct value and the actual value of each data element in the information product [64]. The timeliness [10] dimension of data quality is defined as the currency and volatility of data where currency refers to the age of data and volatility refers to the period of validity for the data.

5.2 Automatic Weighting of Data Attributes

Given a database, a *data attribute* is an attribute that describes the data whereas a *data element* is the value of an attribute associated to a particular data entry. In the case of this study, the data attributes for the subject pre-registration of FICT include Student ID, Student Name, Course Code, Academic Year, Trimester and others, whereas a data element is the value of each attribute for a specific student. In IP-MAP [64], a *simple component* is a combination of the raw data elements for one particular data attribute while an *intermediate component* is a combination of the simple components and data elements. The following formulation is applied to compute the completeness dimension of data and combine them into a single value to be presented to the management users:

$$C^{SC}(c) = \sum_{b=1}^{m} w_b * C^D(b) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(1)}$$

$$C^{IC}(k) = \sum_{c=1}^{n} w_c * C^{SC}(c) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2)}$$

subject to

$$\sum_b w_b = 1 \text{ and } \sum_c w_c = 1$$

where m is the number of raw data element, n is the number of simple component, $C^D(b)$ stands for the completeness of a raw data element b, $C^{SC}(c)$ stands for the completeness of a simple component c, $C^{IC}(k)$ stands for the completeness of an intermediate component k, w_b or w_c is a weighting value between 0 and 1 assigned by users to specify the importance or relevance of the b^{th} raw data element or c^{th} simple component. The same formulation can be directly applied to the accuracy dimension as well, and its description will not be repeated here.

The problem with Equation (1) & (2) is that there is no means of computing w_b and w_c. In reality, these weights need to be manually assigned by the users but this becomes infeasible when the number of attributes or data entries become huge. This is because it will be time consuming and inconvenient for the users to enter the weight of data one by one. Therefore, in practice, uniform weighting is applied but clearly, different attributes have different degree of impact towards decision making.

To alleviate this situation, we propose a novel approach to perform automatic assignment of the attribute weight, w_c to help users identify the attributes that will severely impact the data quality of an information product. In the proposed approach, the weights to the data attribute will be assigned automatically based on its types of usage in the BI Dashboard. This approach will work most of the time because a consistent value of data weight can be generated for each data attribute to determine its importance towards different cases of decision making.

Denote the set of all information processes in BIP-MAP Layer 2 as **I** and the set of all information products, or equivalently dashboards as **D**. The weights are assigned based on its type of usage in the BI Dashboard as follows:

$$w_c = df(c).ipf(c) \quad (3)$$

$$df(c) = \sum_{d=1}^{|\mathbf{D}|} w_{usage}(c,d) \big/ |\mathbf{D}| \quad (4)$$

$$ipf(c) = log(\frac{|\mathbf{I}|}{|\{i:c\in i,\ \ i\in\mathbf{I}\}|}) \quad (5)$$

where $|\cdot|$ is used to denote the cardinality of a set, w_{usage} (*c, d*) is the weight assigned to data attribute *c* based on its usage in the dashboard *d*, $\{i : c \in i,$ $i \in \mathbf{I}\}$ is the set of information processes that contain the data attribute *c*. In summary, *df(c)* is the weighted frequency of the attribute *c* of all dashboards in **D**, *ipf(c)* is the inverse frequency of *c* in all the processes being modeled in the information process layer (BIP-MAP Layer 2).

Equations (3) to (5) are similar to the *Term Frequency and Inverse Document Frequency* (tf-idf) method [28]. tf-idf is a popular and effective method to identify the importance of a word to a document in a collection of documents for information retrieval. In essence, the value of *tf*, analogous to *df* in Equation (4), increases if the keyword can be found in a particular document whereas *idf*, analogous to *ipf* in Equation (5), penalizes the weight if the keyword is also present at most documents in the entire collection.

Figure 5.1 shows an example of the Structure Query Language (SQL) used to generate one of the dashboards. The weight of usage w_{usage} for a data attribute is determined by its function (*condition*, *data*, *information* or any combination of them) in the query process to generate the dashboard, where w_{usage} is incremented with the following values if it is used for any of these functions:

Data Usage	Weight
Condition	0.1
Data	0.4
Information	0.5

Example of SQL

```
SELECT ExamSession, Course, UnitCode, SUM(ActualStudentNo)
AS ActualStudentNo, SUM(ProjectStudentNo) AS ProjectStudentNo
FROM FactStudentFigure
WHERE SUBSTRING(UnitCode,2,1) = 'C'
AND Course = @StudentCourse
AND ExamSession = @ExamSession
AND AcademicYear = @AcademicYear
GROUP BY ExamSession, Course, UnitCode
```

Figure 5.1: An Example of SQL Query for Generating the Dashboard. The query imposes some *condition* (UnitCode, Course, ExamSession, AcademicYear) to extract some *information* (ActualStudentNo and ProjectStudentNo) which will be ordered based on some specified *groups* (ExamSession, Course, UnitCode).

The weight for the information function is assigned a higher value because it is the actual details requested by the users for decision making. The data function should be assigned a higher weight than the condition function because these data attributes categorize the important information into appropriate groups that can be easily identified by the users for performing data analysis. Last but not least, the condition function is assigned a lower weight because these data attributes work at the background to retrieve the relevant data and may not be visible to the users.

The weight of data usage for each function (Condition, Data and Information) is determined empirically by conducting some experiments. First, ten faculty management committee members from FICT in UTAR are invited to determine the importance of each data attribute by providing a weight for each of them. Then, a script is coded to generate all the combinations of weight for data usage that are valid in which their values are summed up to 1. Each combination of weight for data usage is utilized to calculate different sets of weight for all the data attributes. After that, each set of data weight is compared to the average data weight assigned by the ten members. Figure 5.2 shows a surface graph that displays the experiment results of data weight using different set of weights for data usage. It indicates the difference between the data weights assigned by users and the data weights generated by BIP-MAP. The combination of weight for data usage (Condition = 0.1, Data = 0.4, Information = 0.5) that generated a set of data weight that produced the least difference is selected to be implemented into the automatic weighting algorithm for the data attributes.

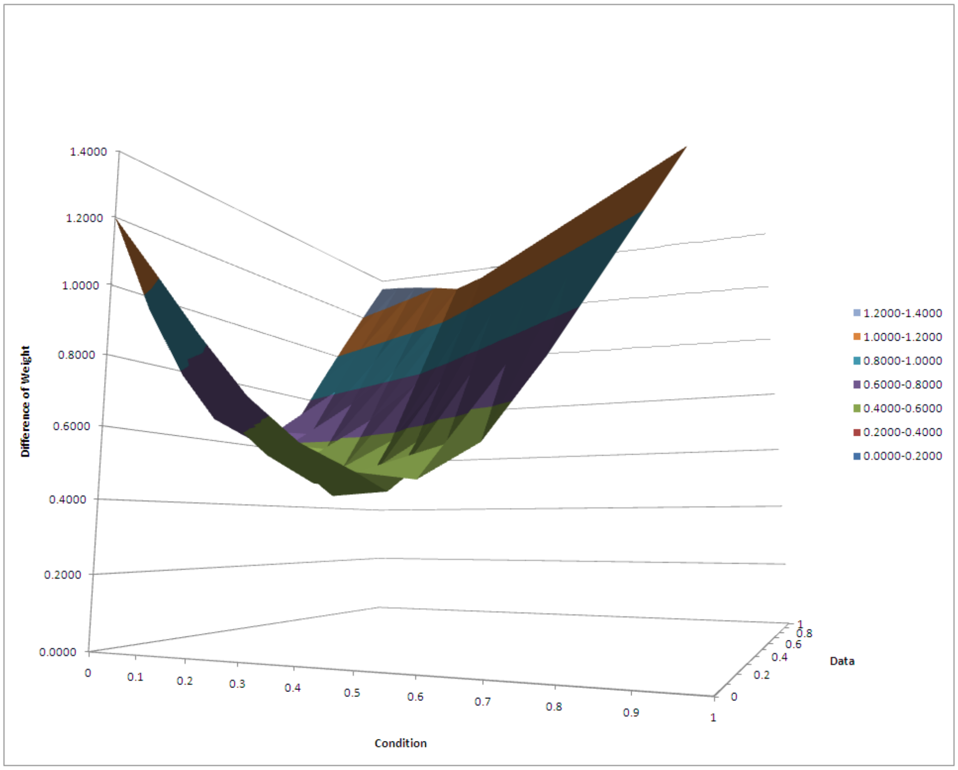


Figure 5.2: Difference of Data Weight between User Assigned Weighting and BIP-MAP Weighting. When the Condition function is weighted as 0.1, the Data function is weighted as 0.4 and the Information function is weighted as 0.5, BIP-MAP produced the least difference of data weight compared to the User Assigned Weighting.

Table 5.1 shows the sample data for computing the completeness data quality dimension with *Uniform Weighting*, *User Assigned Weighting* and *BIP-MAP Weighting*. Each data attribute is considered as equally important or relevant for a decision making in Uniform Weighting. The User Assigned Weighting enables users to assign a weight to each data attribute because the data attributes are not considered as equally important or relevant for a decision making. For the BIP-MAP Weighting, the importance of each data attributes towards decision making is determined automatically based on their usage in the BI Dashboard. If the data is missing, $C^D(b)$ = 0, else $C^D(b)$ = 1. The completeness of a data set is calculated in the same way using the three different types of data weighting methods. The following example computes the completeness of the entire sample data with the BIP-MAP Weighting.

Table 5.1: Computation of Data Quality with Different Weighting Methods

Data Attribute	**Exam Session**	**Course Code**	**Academic Year**	**Trimester**	**UnitCode**	**Unit Description**	**Actual StudentNo**	**Project StudentNo**
Sample Data	201105	CS	1	1	UCCD 1004	PCP	36	100
	NULL	IA	1	1	UCCD 1023	DSA	NULL	26
	201105	IB	1	1	UCCN 1003	DCN	NULL	26
	201105	CN	1	1	MPW 2113	BK	27	NULL
Data Weight								
User Assigned Weighting	0.0870	0.0700	0.0805	0.0805	0.1635	0.1180	0.2175	0.1830
Uniform Weighting	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250	0.1250
BIP-MAP Weighting	* **0.0929**	0.1563	* **0.0715**	* **0.0511**	0.0000	* **0.1226**	* **0.3086**	* **0.1970**
Data Quality (Completeness)								
User Assigned Completeness	0.0653	0.0700	0.0805	0.0805	0.1635	0.1180	0.1088	0.1373
Uniform Completeness	0.0938	0.1250	0.1250	0.1250	0.1250	0.1250	0.0625	0.0938
BIP-MAP Completeness	* **0.0697**	0.1563	* **0.0715**	* **0.0511**	0.0000	* **0.1226**	* **0.1543**	* **0.1478**

$C^{SC}(c) = (1 + 0 + 1 + 1) / 4 \text{ x } 0.0929$
$= 0.0697$ (ExamSession)

$C^{SC}(c) = (1 + 1 + 1 + 1) / 4 \text{ x } 0.1563$
$= 0.1563$ (CourseCode)

$C^{SC}(c) = (1 + 1 + 1 + 1) / 4 \text{ x } 0.0715$
$= 0.0715$ (AcademicYear)

$C^{SC}(c) = (1 + 1 + 1 + 1) / 4 \text{ x } 0.0511$
$= 0.0511$ (Trimester)

$C^{SC}(c) = (1 + 1 + 1 + 1) / 4 \text{ x } 0.0000$
$= 0.0000$ (UnitCode)

$C^{SC}(c) = (1 + 1 + 1 + 1) / 4 \text{ x } 0.1226$
$= 0.1226$ (UnitDescription)

$C^{SC}(c) = (1 + 0 + 0 + 1) / 4 \text{ x } 0.3086$
$= 0.1543$ (ActualStudentNo)

$C^{SC}(c) = (1 + 1 + 1 + 0) / 4 \text{ x } 0.1970$
$= 0.1478$ (ProjectStudentNo)

$C^{IC}(k) = 0.0697 + 0.1563 + 0.0715 + 0.0511 + 0.1226 + 0.1543 + 0.1478$
$= \mathbf{0.7733}$

From the computation results of data quality in Table 5.1, three different set of completeness values are obtained for the sample data. For the Uniform Weighting method, the completeness value of the sample data is 0.8751 and its value will be the same even though the contents of different data attributes are missing as long as the number of missing data is the same. This is not a practical method to calculate the completeness of data because each data attribute has a different degree of impact towards decision making. The method of User Assigned Weighting produced a completeness value of 0.8239 for the sample data because it allows users to allocate different weights to each data attribute in order to specify their importance towards a specific case of decision making. However, this method is infeasible because it will be time consuming and inconvenient to the users when the number of data attributes are large. The proposed method of BIP-MAP Weighting generated a completeness value of 0.7733 for the sample data where the weight of each data attribute is automatically allocated by the algorithm of the framework. According to Table 5.1, most of the data weights generated by BIP-MAP are having a closer value to the data weights assigned by users compared to the uniform weighting. The BIP-MAP generated data weights that are having a closer value to the data weights assigned by users are bolded and highlighted with an asterisk respectively. This indicated that BIP-MAP is able to compute more realistic values of completeness for the sample data compared to the values calculated with uniform weighting because the weight of each data attribute is automatically generated by BIP-MAP based on its function or usage in the BI Dashboard.

To introduce more flexibility, users are also given the option to overwrite the system generated data weight if it is not applicable to the case of their decision making. Figure 5.3 shows a screenshot of the Update Data Weight Module. The first column shows the data attributes and the second column displays the data weight generated by BIP-MAP. Users can enter their preferred data weights to the textboxes at the third column and save them into the database.

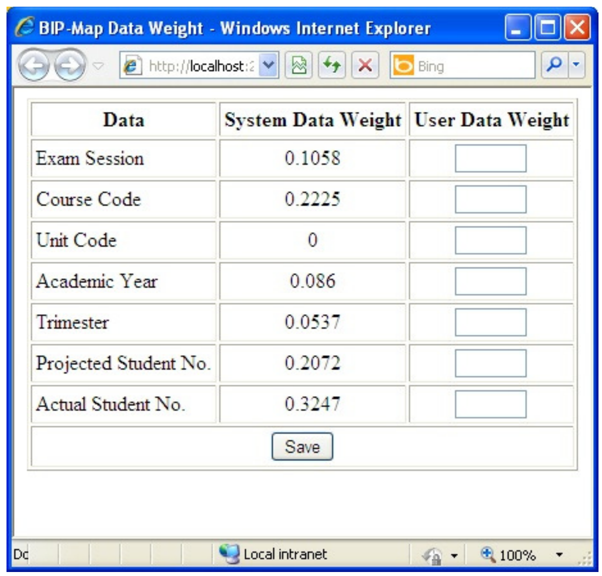


Figure 5.3: Update Data Weight Module. Users are allowed to enter their preferred weight for the data if they feel that the system generated weight is not applicable.

5.3 Consistent Time Management for Business and Information Processes

Users at the management level are usually interested to estimate the time required for generating a set of information reports. Figure 5.5 shows a prototype of the Project Evaluation & Review Technique (PERT) function in the BIP-MAP framework. The PERT timing analysis has been selected to perform time management. Given the duration time for each process, PERT will return some parameters like *early finish time*, *late finish time, slack time* and the *critical path*. The early finish time is the earliest time that a business process or information process can be completed whereas the late finish time is the latest time that a business process or information process must be completed. The slack time is the amount of time that a business process or information process can be delayed. Finally, critical path is the path that consists of all processes that cannot be delayed.

The challenge for implementing PERT for BIP-MAP is to ensure consistencies in the timing analysis between the two layers. To address this issue, for business processes (BIP-MAP Layer 1) that can be mapped to some set of information processes, their input parameters (duration required to complete the process) will be derived from the PERT timing analysis at the information process layer (BIP-MAP Layer 2). The reason is because the duration time of a business process is dependent on the total duration time of all the relevant information processes.

Denote the business process in BIP-MAP Layer 1 as $\mathbf{B} = \{b_1, b_2, b_3, \ldots\}$ and the related information process in BIP-Map Layer 2 as $\mathbf{I} = \{i_1, i_2, i_3, \ldots\}$. There is an one-to-many mapping between $\mathbf{B}$ and $\mathbf{I}$, or $b_i \rightarrow \mathbf{I}(b_i)$ where $\mathbf{I}(b_i) \subseteq \mathbf{I}$ is the set of information processes mapped by the business process b_i. Algorithm 1 shows the adaptation of PERT time management algorithm for BIP-MAP. The business processes in BIP-MAP Layer 1 may or may not be mapped to the information processes in BIP-MAP Layer 2. For the latter, users specify the duration input directly at the business process layer. For the former, users input the duration at the information process layer. The

PERT timing analysis is then conducted on the mapped information processes $\mathbf{I}(b_i)$ and the result is then propagated to b_i as the duration time parameter.

If an optimistic view is preferred, the maximum early finish time of the information processes $\mathbf{I}(b_i)$ will be propagated to generate the *optimistic duration time* of the business process b_i as follows:

$$do^{BP}(b_i) = \max\left(\mathbf{e}^{IP}\left(\mathbf{I}(b_i)\right)\right) \quad \ldots\ldots(7)$$

where $do^{BP}(b_i)$ is the optimistic duration time of the business process b_i and $\boldsymbol{e}^{IP}\left(\mathbf{I}(b_i)\right)$ is a set of early finish times of all the information processes $\mathbf{I}(b_i)$ mapped by the business process b_i.

On the other hand, if a pessimistic view is preferred, the maximum late finish time of the information processes $\mathbf{I}(b_i)$ will be propagated to generate the *pessimistic duration time* of the business process b_i as follows:

$$dp^{BP}(b_i) = \max\left(\mathbf{l}^{IP}\left(\mathbf{I}(b_i)\right)\right) \quad \ldots\ldots(8)$$

where $dp^{BP}(b_i)$ is the pessimistic duration time of the business process b_i and $\boldsymbol{l}^{IP}\left(\mathbf{I}(b_i)\right)$ is a set of late finish times of all the information processes $\mathbf{I}(b_i)$ mapped by the business process b_i.

<u>Algorithm 1: Time Management</u>

Input:

$d^{BP}(b_i), \mathbf{I}(b_i) = \emptyset$: The duration time of a business process that is not mapped to any information process.

$d^{IP}(i_i)$: The duration time of an information process.

Output:

$e^{BP}(b_i)$: The early finish time of a business process.

$l^{BP}(b_i)$: The late finish time of a business process.

$s^{BP}(b_i)$: The slack time of a business process.

Method:

for b_1 to b_n **do**

    if $\mathbf{I}(b_i) = \emptyset$ **then**

        Capture $d^{BP}(b_i)$ from the users

    else

        Capture $d^{IP}(i_i), \forall\, i \in \mathbf{I}(b_i)$ from the users

        Perform PERT calculation on $\mathbf{I}(b_i)$

        if optimistic time is preferred **then**

            Propagate $do^{BP}(b_i)$ from layer 2 to layer 1 with Equation (7)

        else

            Propagate $dp^{BP}(b_i)$ from layer 2 to layer 1 with Equation (8)

        end if

    end if

end for

Perform PERT calculation to retrieve each $e^{BP}(b_i)$, $l^{BP}(b_i)$, $s^{BP}(b_i)$ and the critical path

Figure 5.4: Time Management Algorithm

Figure 5.5 shows an example of how the duration time for the business process of *Prepare Projected No. of Students* is derived from the early/late finish time of the corresponding information processes in the second layer. After all the optimistic duration time or pessimistic duration time of each business process b_i is derived from the information process layer, the parameters like early finish time, late finish time, slack time and the critical path of all the business processes can be computed using the PERT calculation methods. The results of the PERT calculation are provided in Table 5.2.

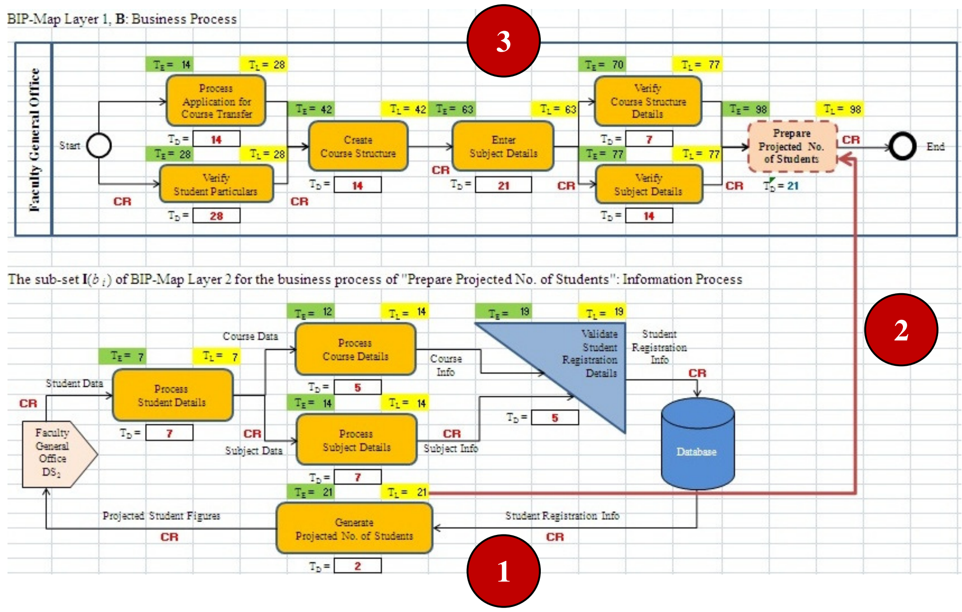


Figure 5.5: PERT Function for Time Management. (1) At each sub-set of BIP-MAP Layer 2, PERT is performed on to generate the early finish time (green) and late finish time (yellow) of each information process. (2) The maximum early/late finish time of is propagated to b_i as the duration input. (3) Then, PERT is performed at BIP-MAP Layer 1 to generate the early finish time (green) and late finish time (yellow) of each business process.

Table 5.2: PERT Calculation Results

Business Process	Time (Day)				
	Duration	**Early Finish**	**Late Finish**	**Slack**	**Critical Path**
Process Application for Course Transfer	14	14	28	14	No
Verify Student Particulars	28	28	28	0	Yes
Create Course Structure	14	42	42	0	Yes
Enter Subject Details	21	63	63	0	Yes
Verify Course Structure Details	7	70	77	7	No
Verify Subject Details	14	77	77	0	Yes
Prepare Projected No. of Students	21	98	98	0	Yes
Information Process	**Time (Day)**				
	Duration	**Early Finish**	**Late Finish**	**Slack**	**Critical Path**
Process Student Details	7	7	7	0	Yes
Process Course Details	5	12	14	2	No
Process Subject Details	7	14	14	0	Yes
Validate Student Registration Details	5	19	19	0	Yes
Generate Projected No. of Students	2	21	21	0	Yes

5.4 Evaluation and Discussion

To confirm the usefulness of the implementation for data quality and time management, BIP-MAP is compared against the original BPMN and IP-MAP models for the subject pre-registration of FICT in UTAR. The evaluation is performed by the same ten members from the faculty management committee for the three models using the five-point likert scale respectively. An overall system performance is conducted for the three models by allowing the evaluators to rate and rank each feature in a comparative manner. Likewise, the evaluators provided their additional feedbacks by answering some general questions about the BIP-MAP framework. Table 5.3 provides a brief summary of the features that are evaluated for the three models with an average rating from all the ten evaluators.

5.4.1 Time Management

All the three models are rated with a high value that is more than 4 for the time management feature. This showed that it is useful to identify the timeliness details so that users are able to estimate the time required to complete each process in a business operation. For example, the timeliness details that can be generated are the early finish or late finish time, slack time and critical path. The definitions of these timeliness details are provided in Section 5.3.

Table 5.3: Features Evaluated in BPMN, IP-MAP and BIP-MAP

BPMN		
Feature	**Usefulness**	**Availability in IP-MAP**
1. Time Management Enable users to plan the duration time for each activity of a business process when it is integrated with the PERT chart.	✓ (4.4)	✓ (4.33)
IP-MAP		
Feature	**Usefulness**	**Availability in BPMN**
1. Time Management Enable users to estimate the time required for generating a set of information reports by determining the critical path.	✓ (4.4)	○ (3.89)
2. Data Quality Management Allows users to assign different weights to specify the importance or relevance of the data towards a case of decision making.	✗ (2.8)	✗ (1.78)
BIP-MAP		
Feature	**Usefulness**	
1. Time Management Timing parameters from the information process layer are automatically transformed into the business process layer for users to perform time management.	✓ (4.1)	
2. Data Quality Management A weight is automatically calculated for the data based on the frequency and condition of its usage in the BI Dashboard.	○ (3.5)	

Symbol	**Range of Average Rating**
✓	4.00 – 5.00 (Agree / Strongly Agree)
○	3.00 – 3.99 (Neutral)
✗	1.00 – 2.99 (Strongly Disagree / Disagree)

When BPMN is compared to IP-MAP, the time management feature is rated with a moderate value of 3.89 only because the feature of time management is poorly addressed in the original BPMN model [78]. Nevertheless, the evaluation results of 4.4 for BPMN shows that time management is recognized as a useful feature if the PERT calculation technique is being integrated into it.

IP-MAP is rated with a high value of 4.4 and 4.33 respectively because the PERT calculation technique is available in the original IP-MAP model [63] for users to estimate the time required to generate the information products at different stages of the information manufacturing chain. This phenomenon happened because the evaluators considered the time management feature in IP-MAP to be more comprehensive at the implementation level of information processes compared to BPMN. However, integrating PERT into BPMN will provide the users a more general view at the management level for the business processes.

BIP-MAP received a high rating of 4.1 for the time management feature. This indicated that almost every evaluator either agreed or strongly agreed on the usefulness of propagating the timing parameters from the second layer of BIP-MAP (information process layer) to its first layer (business process layer). When the timing parameters are propagated from the information process layer to the business process layer, users are able to estimate the time that is required for generating each BI product in a consistent manner. A detailed description of the time management feature for BIP-MAP is provided in Section 5.3.

5.4.2 Data Quality Management

For the data quality feature, BPMN is rated with a very low rating of 1.78 because it focuses on modeling the business processes only and the information processes that describe how data is being manipulated are not modeled. Even though IP-MAP models the information processes that are involved in generating the information products of an organization, the data quality feature is rated with a low rating of 2.8 for several reasons. One of the evaluators felt that using the method of weight assignment to specify the importance or relevance of data in computing the data quality is arguable since different users may assign different weights for the same set of data. The range of data quality values that should be acceptable to the users is also questioned by one of them. Two evaluators indicated that the formula for computing data quality is good but they do not considered it as useful because they felt that the data should be made available by the system if it is required by the users.

Although the method of weight assignment seems to be arguable, it enables users to acquire a more relevant data quality value that describes the completeness or accuracy of data compared to the method of uniform weighting. Even though different weights are assigned by different users, the same important data attributes are identified by them. The range of data quality values that are acceptable to the users are determined by different cases. Hence, the range of data quality values that are acceptable to the users cannot be specified in a standardized manner. It is risky to always think that the systems should be providing the complete or accurate set of data for information reporting as system errors may happen and cause the occurrence of incomplete or inaccurate data. Thus, it is absolutely important for the management users to be assured of the quality of data before using it for any decision making.

Even though the data quality feature of BIP-MAP received a moderate rating of 3.5, its degree of usefulness is still higher compared to the original IP-MAP due to the implementation of automatic weighting where the weight of each data attribute is determined by its function in the query processes used to generate the BI products. With the implementation of automatic weighting, the weight of each data attribute can be defined automatically in a more relevant manner and users are not required to spend time in performing the manual task of weight assignment for all the data attributes. A detailed description of the automatic weighting feature for BIP-MAP is provided in Section 5.2.

5.4.3 Overall Performance

As the data quality and time management features of BPMN, IP-MAP and BIP-MAP are compared and evaluated together, a score between 0 to 5 is provided by the evaluators to rate the overall performance of the three models and ranked them accordingly. The results of evaluation for the three models are shown in Figure 5.6. It shows that BIP-MAP outperforms BPMN and IP-MAP as it receives a higher score or rank for both the features of data quality and time management among the three models. In general, almost all of the evaluators provided a high score for the time management feature because users are able to estimate the time that is required to complete each activity available in a process. More than half of them supported the feature of automatic data weighting with a high scoring because it helps users to identify the importance of data by providing a weight for each data item.

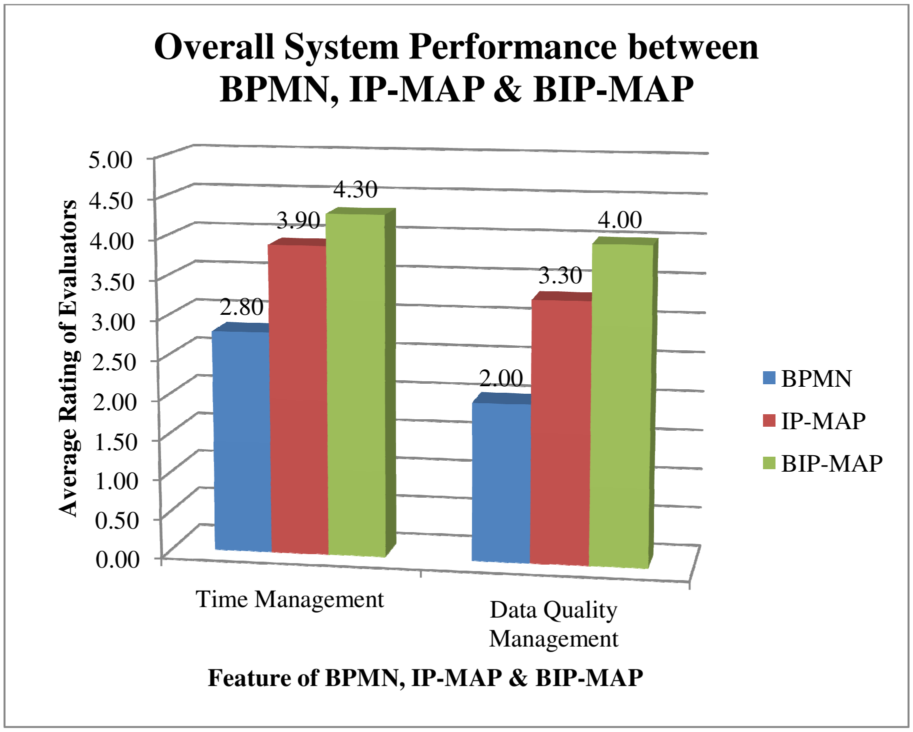


Figure 5.6: Overall System Performance between BPMN, IP-MAP and BIP-MAP (proposed framework). The features of data quality and time management in the three models are rated by the evaluators in a comparative manner.

The evaluators also gave the following feedback to further enhance BIP-MAP:

- **Implementing Process Monitoring for Time Management**
  - ✓ One of the evaluators considered that BIP-MAP is good for the planning stage but has unsure performance in implementation because it is not designed to measure the performance of execution for the business or information processes. The evaluator felt that using BIP-MAP for only planning in time management is not sufficient and the performance of execution for each process should be monitored in order to ensure a successful project management. The time management feature of BIP-MAP can be enhanced by adding a monitoring function that is integrated with the system clock in order to

allow users to keep track of the execution of each business or information process. With a monitoring function, users will be able to ensure that each process is performing as scheduled.

- ✓ Another evaluator indicated that the timing parameters available in BIP-MAP are too many but the only important timing information that is required is the milestone or deadline of each process. The timing parameters in BIP-MAP can be reorganized in order to enable users to have a clearer view of the timeliness details for each process. Apart from this, the evaluator proposed to integrate BIP-MAP with the system clock so that users can identify the progression of each process in line with the actual time. Indeed, BIP-MAP can be integrated with the system clock in order to implement a real time monitoring for both the business and information processes.

- **Enhancing the Automatic Weighting for Data Attributes**
  - ✓ One of the evaluators pointed out that the weight of usage for each data attribute in the query processes should be estimated based on a set of configuration where users are allowed to specify its level of importance towards a specific case even though automatic weighting is implemented. One of the examples of configuration given by the evaluator is to use the five-point likert scale (1 – very unimportant to 5 – very important). A configuration function which allows users to update the weight of usage for each function of data attribute in the query processes can be implemented so that the relevant data weights can be generated for different cases.

- ✓ Another evaluator still considered that using automatic weighting to specify the importance of each data attribute is arguable because the weight of usage for each function of data attribute in the query processes can be specified with a different value by different users. Therefore, automatic weighting may not be able to identify the actual level of importance for each data attribute towards to a specific case. Even though the method of automatic weighting still seems to be arguable, it has reduced the job of manual weight assignment and provided users a guideline to comprehend the quality of data before using the data for any decision making.

5.5 Chapter Summary

This chapter has presented two algorithms in BIP-MAP. The first algorithm implements a function to perform automatic weighting for determining the importance of each data attribute in a system so that the quality of data can be computed in a more realistic manner. The second algorithm integrates the PERT function to implement a consistent time management for both layers of the business and information processes where users are able to estimate various timing parameters at different stages of a business or information process. It is confirmed through the case study that BIP-MAP outperformed the original BPMN and IP-MAP for these two features.

CHAPTER 6

CONCLUSION

6.1 Summary of Contribution

Three major contributions have been achieved in this research. First, a framework that implements BPM and TDQM for a BI platform is developed by integrating the salient modeling and management techniques of BPMN and IP-MAP. By integrating BPMN and IP-MAP, the BI products that are generated at the information process layer can be summarized onto the business process layer so that management users are able to easily identify all the BI products that are available to be utilized for business decision making. As BPMN and IP-MAP are integrated into a single framework, users are able to identify how the data is being captured, validated, processed, stored, transformed and generated throughout the organization by drilling down into the detailed data processing steps that are involved in each business process. Apart from this, the integration in BIP-MAP enables users to have a more complete understanding of all the business and information system operations because the technical users are able to understand the business considerations behind certain data, while the business users can get an insight into the logic behind the information generated to them.

Second, a novel approach of automatic weight assignment for the data attributes is produced to help users identify the importance of each data attribute in order to determine the actual data quality of a BI product. As the weight of each data attribute is calculated automatically based on their usage in the BI Dashboard, a more realistic and consistent value for the data quality dimensions of completeness and accuracy can be computed for the BI products. In addition, users are not required to spend their precious time in assigning weights to the data attributes even if the data set contains a large number of attributes.

Last but not least, an integrated time management feature is implemented to assist users in performing planning between the two layers of business and information processes. With the implementation of this feature, consistencies for the timing parameters between the two layers are achieved because users are only required to input the timing parameters at the information process layer for the business processes that are mapped to a set of information processes. The total duration time of the information processes will be automatically propagated up to the relevant process at the business process layer. As the duration time for each process is determined at both layers, the PERT algorithm can be executed to compute various timeliness details which can assist the users to estimate the time that is required for accomplishing a task or generating a report at different phases of a process.

6.2 Future Research for BIP-MAP

After conducting an evaluation for BIP-MAP with the management users of the university and the online job recruitment company, some limitations of the framework are identified for future research. The major concern of the management users for the framework is the method of creating and updating the BIP-MAP for complex systems especially those that have ongoing changes. This is because the time spent in producing the BIP-MAP may be quite long if it is being constructed manually. However, this should not be a major problem because it can be resolved by developing software with user-friendly interface to assist management users in creating and updating the BIP-MAP. The software can be developed with a drag and drop development approach so that the BIP-MAP can be constructed in a fast and easy manner. In addition, mechanism to associate metadata to each element should also be included for enabling users to model the business and information processes and manage the metadata with ease.

Apart from this, some minor limitations are also identified for the framework. BIP-MAP can be enhanced by including a function to provide users some possible options of solution for any problem that may happen to the BI products. Additional information that describes the details of any document can also be included at the business process layer so that users are able to identify important information for all the documents that are available in a business process. For the search function of the BI Dashboard, it can be enhanced by implementing a "What-If" analysis feature to enable users to view the data in different perspectives. The time management feature can be improved by adding a monitoring function that is integrated with the system clock so that users are able to keep track of the execution of all processes. Last but not least, the access control feature can be added in order to enforce information security by delivering the right information to the right users.

REFERENCES

[1] OMG, (2011) "OMG UML Superstructure Version 2.4.1," [Online] Available:

[2] M. A. Ould, *Business Processes: Modelling and Analysis for Re-engineering and Improvement*. New Jersey: Wiley, 1995.

[3] S. A. White, (2004) "Introduction to BPMN," IBM Corporation [Online] Available: http://www.omg.org/bpmn/Documents/Introduction_to_BPMN.pdf

[4] I. Jacobson, G. Booch, and J. Rumbaugh, *The Unified Software Development Process*. New Jersey: Addison Wesley Professional, 1999.

[5] R. J. Mayer, M. K. Painter, and P. S. deWitte, (1992) "IDEF Family of Methods for Concurrent Engineering and Business Re-engineering Applications," [Online] Available: http://www.iso.staratel.com/IDEF/BPR/IDEFFAMI.pdf

[6] K. Jensen, "A Brief Introduction to Coloured Petri Nets," presented at the *TACAS '97 Proceedings of the Third International Workshop on Tools and Algorithms for Construction and Analysis of Systems*, Enschede, The Netherlands, 2-4 April 1997.

[7] G. Shankaranarayanan, R. Y. Wang, and M. Ziad, "IP-MAP: Representing the Manufacture of an Information Product," presented at the *Proceedings of the 2000 International Conference on Information Quality*, MIT, 20-22 October 2000.

[8] G. Keller, M. Nüttgens, and A.-W. Scheer, (1992) "Semantische Prozessmodellierung auf der Grundlage Ereignisgesteuerter Prozessketten (EPK)," Universität des Saarlandes [Online] Available: http://www.wiso.uni-hamburg.de/fileadmin/wiso_fs_wi/Team/Mitarbeiter/Prof._Dr._Markus_Nuettgens/Publikationen/heft089.pdf

[9] M. Scannapieco, B. Pernici, and E. Pierce, "IP-UML: Towards a Methodology for Quality Improvement Based on the IP-MAP Framework," presented at the *Proceedings of the Seventh International Conference on Information Quality (ICIQ-02)*, MIT, 8-10 November 2002.

[10] D. Ballou, R. Wang, H. Pazer, and G. Kumar.Tayi, "Modeling Information Manufacturing Systems to Determine Information Product Quality," *Management Science,* vol. 44, pp. 462-484, April 1998.

[11] Gartner, (2009) "Gartner EXP Worldwide Survey of More than 1,500 CIOs Shows IT Spending to Be Flat in 2009," Gartner [Online] Available: http://www.gartner.com/it/page.jsp?id=855612

[12] Gartner, (2011) "Gartner Says Worldwide Business Intelligence, Analytics and Performance Management Software Market Surpassed the $10 Billion Mark in 2010," Gartner [Online] Available: http://www.gartner.com/it/page.jsp?id=1642714

[13] B. Wixom and H. Watson, "The BI-Based Organization," *International Journal of Business Intelligence Research,* vol. 1, pp. 13-28, January-March 2010.

[14] G. Lawton, "Making Business Intelligence More Useful," *IEEE Computer,* vol. 39, pp. 14-16, September 2006.

[15] J. A. Hoffer, J. F. George, and J. S. Valacich, *Modern Systems Analysis And Design*. New Jersey: Pearson Education, 2008.

[16] E. Turban, R. Sharda, J. E. Aronson, and D. King, *Business Intelligence: A Managerial Approach*. New Jersey: Pearson Education, 2007.

[17] C. M. Olszak and E. Ziemba, "Approach to Building and Implementing Business Intelligence Systems," *Interdisciplinary Journal of Information, Knowledge, and Management,* vol. 2, pp. 135-148, 2007.

[18] L. Xu, L. Zeng, Z. Shi, Q. He, and M. Wang, "Research on Business Intelligence in Enterprise Computing Environment," presented at the *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Montreal, Canada, 7-10 October 2007.

[19] A. Parssian, "Managerial Decision Support with Knowledge of Accuracy and Completeness of Relational Aggregate Functions," *Decision Support Systems,* vol. 42, pp. 1494-1502, December 2006.

[20] T. H. Davenport, "Business Intelligence and Organizational Decisions," *International Journal of Business Intelligence Research,* vol. 1, pp. 1-12, January-March 2010.

[21] E. Thomsen, (2003) "BI's Promised Land," Intelligent Enterprise [Online] Available: http://www.providersedge.com/docs/km_articles/BI-s_Promised_Land.pdf

[22] S. Chaudhuri, U. Dayal, and V. Narasayya, "An Overview of Business Intelligence Technology," *Communications of the ACM,* vol. 54, pp. 88-98, August 2011.

[23] R. S. Aguilar-Saven, "Business Process Modeling: Review and Framework," *International Journal of Production Economics,* vol. 90, pp. 129-149, 2003.

[24] A. Groznik and A. Kovacic, "Business Renovation: From Business Process Modelling to Information System Modelling," presented at the *24th International Conference of Information Technology Interfaces*, Cavtat, Croatia, 24-27 June 2002.

[25] R. Y. Wang, Y. W. Lee, L. L. Pipino, and D. M. Strong, "Manage Your Information as a Product," *Sloan Management Review,* vol. 39, pp. 95-105, 1998.

[26] G. Shankaranarayanan and A. Even, "The Metadata Enigma," *Communications of the ACM,* vol. 49, pp. 88-94, February 2006.

[27] G. Shankaranarayanan, "Towards Implementing Total Data Quality Management In A Data Warehouse," *Journal of Information Technology Management,* vol. 16, pp. 21-30, 2005.

[28] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries," presented at the *The First Instructional Conference on Machine Learning*, Piscataway, New Jersey, USA, 3-8 December 2003.

[29] T. Chee, L.-K. Chan, M.-H. Chuah, C.-S. Tan, S.-F. Wong, and W. Yeoh, "Business Intelligence Systems: State-Of-The-Art Review And Contemporary Applications," presented at the *Symposium on Progress in Information & Communication Technology*, Kuala Lumpur, Malaysia, 7-8 December 2009.

[30] C.-H. Chee, Y.-W. Sim, and W. Yeoh, "An Integrated Information Manufacturing Framework for Business Intelligence Systems," presented at the *15th IBIMA conference on Knowledge Management and Innovation: A Business Competitive Edge Perspective*, Cairo, Egypt, 6-7 November 2010.

[31] C.-H. Chee, W. Yeoh, and S. Gao, "Enhancing Business Intelligence Traceability through an Integrated Metadata Framework," presented at the *22nd Australasian Conference on Information Systems*, Sydney, Australia, 29 November – 2 December 2011.

[32] E. Simperl, I. Thurlow, P. Warren, F. Dengler, J. Davies, M. Grobelnik, D. Mladeni´c, J. M. Gómez-Pérez, and C. R. Moreno, "Overcoming Information Overload in the Enterprise - The Active Approach," *IEEE Internet Computing,* vol. 14, pp. 39-46, December 2010.

[33] J. C. Hancock and R. Toren, *Practical Business Intelligence with SQL Server 2005*. New Jersey: Addison Wesley Professional, 2006.

[34] D. Loshin, *Business Intelligence: The Savvy Manager's Guide*. San Francisco: Morgan Kaufmann Publishers, 2003.

[35] H. J. Watson and B. H. Wixom, "The Current State of Business Intelligence," *IEEE Computer,* vol. 40, pp. 96-99, September 2007.

[36] S. O. Jr., "Is Business Intelligence a Smart Move?," *IEEE Computer,* vol. 35, pp. 11-14, July 2002.

[37] S. O. Jr., "Taking Business Intelligence to the Masses," *IEEE Computer,* vol. 43, pp. 12-15, July 2010.

[38] S. Viaene, "Linking Business Intelligence into Your Business," *IEEE Computer,* vol. 10, pp. 28-34, December 2008.

[39] E. B. Sloane, E. Rosow, and J. Adam, "Strategic Graphic Dashboards for Improved Technology Management Decisions," presented at the *Proceedings of the 25th Annual International Conference of the IEEE EMBS*, Cancun, Mexico, 17-21 September 2003.

[40] E. B. Sloane, E. Rosow, J. Adam, and D. Shine, "JEDI – An Executive Dashboard and Decision Support System for Lean Global Military Medical Resource and Logistics Management," presented at the *EMBS Annual International Conference*, New York, USA, 30 August - 3 September 2006.

[41] E. Turban, J. E. Aronson, T.-P. Liang, and R. Sharda, *Decision Support and Business Intelligence Systems*. New Jersey: Pearson Education, 2007.

[42] W. v. d. Aalst, "Using Process Mining to Bridge the Gap between BI and BPM," *IEEE Computer,* vol. 44, pp. 77-80, December 2011.

[43] H. A. Reijers and J. Mendling, "A Study Into the Factors That Influence the Understandability of Business Process Models," *IEEE Transactions On Systems, Man, And Cybernetics - Part A: Systems And Humans,* vol. 41, pp. 449-462, May 2011.

[44] M. Weske, *Business Process Management: Concepts, Languages, Architectures*. Berlin: Springer, 2007.

[45] W. Tan, W. Shen, L. Xu, B. Zhou, and L. Li, "A Business Process Intelligence System for Enterprise Process Performance Management," *IEEE Transactions On Systems, Man, And Cybernetics - Part C: Applications And Reviews,* vol. 38, pp. 745-756, November 2008.
[46] L. Aldin and S. d. Cesare, "A Comparative Analysis Of Business Process Modelling Techniques," presented at the *Proceedings of the U.K. Academy for Information Systems (UKAIS 2009)*, Oxford, UK, 31 March - 1 April 2009.
[47] H. Mili, G. Tremblay, G. B. Jaoude, E. Lefebvre, L. Elabed, and G. E. Boussaidi, "Business Process Modeling Languages Sorting Through the Alphabet Soup," *ACM Computing Surveys,* vol. 43, pp. 4:1-56, November 2010.
[48] E. Kindler, "On the semantics of EPCs: A Framework for Resolving the Vicious Circle," August 2003.
[49] I. Reinhartz-Berger, P. Soffer, and A. Sturm, "Extending the Adaptability of Reference Models," *IEEE Transactions On Systems, Man, And Cybernetics - Part A: Systems And Humans,* vol. 40, pp. 1045-1056, September 2010.
[50] J. Soler, I. Boada, F. Prados, and J. Poch, "A Web-based E-learning Tool for UML Class Diagrams," presented at the *IEEE EDUCON Education Engineering 2010*, Madrid, Spain, 14-16 April 2010.
[51] J. Siegel, (2010) "Introduction to OMG's Unified Modeling Language," Object Management Group [Online] Available: http://www.omg.org/gettingstarted/what_is_uml.htm
[52] W. Wang, H. Ding, J. Dong, and C. Ren, "A Comparison of Business Process Modeling Methods," presented at the *IEEE International Conference on Service Operations and Logistics, and Informatics Proceedings*, Shanghai, China, 21-23 June 2006.
[53] L. E. Garcia-Fernandez and M. Garijo, "Modeling Strategic Decisions Using Activity Diagrams to Consider the Contribution of Dynamic Planning in the Profitability of Projects Under Uncertainty," *IEEE Transactions On Engineering Management,* vol. 57, pp. 463-476, July 2010.
[54] S. Han and H. Y. Youn, "Modeling and Analysis of Time-Critical Context-Aware Service Using Extended Interval Timed Colored Petri Nets," *IEEE Transactions On Systems, Man, And Cybernetics - Part A: Systems And Humans,* vol. 42, pp. 630-640, May 2012.
[55] M. Witsch and B. Vogel-Heuser, "Towards a Formal Specification Framework for Manufacturing Execution Systems," *IEEE Transactions On Industrial Informatics,* vol. 8, pp. 311-320, May 2012.
[56] M. Owen and J. Raj, (2003) "BPMN and Business Process Management," Popkin Software [Online] Available: http://www.omg.org/bpmn/Documents/6AD5D16960.BPMN_and_BPM.pdf
[57] OMG, (2009) "BPMN Specifications Version 2.0," [Online] Available:
[58] W. Bandara and M. Rosemann, "What are the Secrets of Successful Process Modeling?," *Systèmes D'Information Et Management,* vol. 10, pp. 47-68, November 2005.
[59] J. Cao, X. Diao, and G. Jiang, "Data Lifecycle Process Model and Quality Improving Framework for TDQM Practices," presented at the

2010 International Conference on E-Product E-Service and E-Entertainment (ICEEE), Henan, China, 7-9 November 2010.

[60] C. Fisher, E. Lauria, S. Chengalur-Smith, and R. Wang, *Introduction to Information Quality*. Bloomington: AuthorHouse, 2012.

[61] S. Sachdeva and S. Bhalla, "Semantic Interoperability in Standardized Electronic Health Record Databases," *ACM Journal of Data and Information Quality,* vol. 3, pp. 1-37, April 2012.

[62] R. Y. Wang, "A Product Perspective on Total Data Quality Management," *Communications of the ACM,* vol. 41, pp. 58-65, February 1998.

[63] G. Shankaranarayanan, M. Ziad, and R. Y. Wang, "Managing Data Quality in Dynamic Decision Environments: An Information Product Approach," *Journal of Database Management,* vol. 14, pp. 14-32, October-December 2003.

[64] G. Shankaranarayanan and Y. Cai, "Supporting Data Quality Management in Decision-Making," *Decision Support Systems,* vol. 42, pp. 302-317, October 2006.

[65] P. Ponniah, *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals*. New York: Wiley, 2001.

[66] A. Sen, "Metadata Management: Past, Present and Future," *Decision Support Systems,* vol. 37, pp. 151-173, December 2002.

[67] Y. Verbitskiy, "An Empirical Study of Metadata Issues in Business Intelligence Environment: An Action Research in a Higher Education Institution.," University of South Australia, 2009.

[68] A. M. Jenkins, "Research Methdologies And MIS Research," in *Research Methods in Information Systems* Amsterdam: Elsevier Science Publishers, 1985, pp. 97-109.

[69] V. Vaishnavi and B. Kuechler, (2011) "Design Science Research in Information Systems," DESRIST [Online] Available: http://desrist.org/design-research-in-information-systems/

[70] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design Science in Information Systems Research," *MIS Quarterly,* vol. 28, pp. 75-105, March 2004.

[71] N. Mack, C. Woodsong, K. M.MacQueen, G. Guest, and E. Namey, *Qualitative Research Methods: A Data Collector's Field Guide*. Research Triangle Park: Family Health International, 2005.

[72] R. K. Yin, *Case Study Research: Design and Methods*, 4th ed. Thousand Oaks: Sage Publications, 2009.

[73] Saperion, (2010) "How to Model BPMN for Dynamic Process Execution," Saperion [Online] Available: http://www.saperionblog.com/lang/de/how-to-model-bpmn-for-dynamic-process-execution/2996/

[74] K. Weber, B. Otto, and H. Osterle, "One Size Does Not Fit All - A Contingency Approach to Data Governance," *Data and Information Quality,* vol. 1, pp. 4:1-27, June 2009.

[75] S. E. Madnick, R. Y. Wang, Y. W. Lee, and H. Zhu, "Overview and Framework for Data and Information Quality Research," *ACM Journal of Data and Information Quality,* vol. 1, pp. 1-22, June 2009.

[76] T. C. Redman, *Data Quality for the Information Age*. Norwood: Artech House, 1997.

[77] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, "Methodologies for Data Quality Assessment and Improvement," *ACM Computing Surveys,* vol. 41, pp. 16:1-52, July 2009.

[78] D. Gagne and A. Trudel, "Time-BPMN," presented at the *IEEE Conference on Commerce and Enterprise Computing*, Montreal, QC, Canada, 20-23 July 2009.

APPENDICES

Appendix A: Survey Questions for Universiti Tunku Abdul Rahman

UNIVERSITI TUNKU ABDUL RAHMAN
Faculty of Information & Communication Technology
Department of Information Systems

An Integrated Framework For Business Intelligence Systems:
Business Intelligence Product Map
By Chin-Hoong Chee

Survey Questions

As a business entity grows in size and profile, it becomes more and more challenging for management users to perform project planning and execution, resource allocation and keep track of the various activities within the organization. Therefore, Business Intelligence (BI) has become an attractive tool to help the management users to handle these complicated tasks.

Generally, BI is a set of concepts, methods, and technologies that play a key role in turning data into information and eventually into knowledge to assist users in performing the business decision making process. In this research, the salient modeling methods of two existing models are investigated and integrated to construct a framework called the Business Intelligence Product Map (BIP-Map). The two models that have been integrated into the framework are:

(a) Business Process Modeling Notation (BPMN)
(b) Information Product Map (IP-Map)

The purpose of this survey is to evaluate the strengths and limitations of each model and the improvements that can be gained by integrating the two models into a single framework. The framework has been implemented to model the *Subject Pre-Registration Process of FICT* in UTAR. Targeted users of the framework include the management users of a faculty like the Dean, Deputy Deans, Head of Departments and Head of Programmes.

Please provide your feedback for each statement with the following ratings.

1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Name of Evaluator: ________________________________

Terminologies

1. **Business Intelligence**
   - A set of concepts, methods, and technologies that play a key role in turning data into information and eventually into knowledge to assist users in performing the business decision making process.

2. **Business Process**
   - A group of activities within an organization that consists of certain logical order and dependence in achieving a particular business objective.

3. **Information Process**
   - A group of activities within the systems of an organization that consists of certain logical order and dependence in processing the data and information.

4. **Information Product**
   - The output of an information process where data and information is treated much like the products that are being manufactured in any industry.

Section A: **Business Process Modeling Notation**	**Rate**	**IP-Map**	**Comments**
BPMN is a model that is developed by the Business Process Management Initiative (BPMI) for users to manage the business processes of their organizations. It provides a standardized notation which is easily understood by all the business users, including users at the management level and technical users who are responsible in developing the program to execute those processes. The statements below describe about the features that are available in BPMN. Please **rate the usefulness of the features** towards business decision making. Please rate the feature at the IP-Map column **if the feature is present in IP-Map**.			
1. BPMN ***breaks down a business process into a set of activities***. Segregation of processes into multiple activities helps the management users to perform decision making for the students and subjects of the faculty.			

2. BPMN helps users ***to visualize the workflow of activities*** within a business process in a clear and proper manner. Therefore, BPMN will enable users ***to identify the appropriate order and dependencies among the activities*** to execute all the activities in the Subject Pre-Registration Process. This helps the users to gain a better understanding of the functional and structural workings of the business process. 3. BPMN allows users ***to easily identify the interaction between all participants*** that are involved in the Subject Pre-Registration Process when the relevant activities to be implemented by each participant are being grouped appropriately into different sections. 4. BPMN allows users ***to perform time management*** where they can plan the duration time ***for each activity of a business process*** when it is integrated with the PERT chart.			
Additional Comments:			
Section B: **Information Product Map**	**Rate**	**BPMN**	**Comments**
It is very important for users to utilize high quality information during their business decision making process. IP-Map is a model used to systematically describe the manufacturing process of an information product in order to implement Total Data Quality Management. In IP-Map, data and information are being treated much like the products that are being manufactured in the industry. The statements below describe about the features that are available in IP-Map. Please **rate the usefulness of the features** towards business decision making. Please rate the feature at the BPMN column **if the feature is present in BPMN**.			

5. The output of IP-MAP includes intermediate and final ***information products***. In the Subject Pre-Registration Process, information products are management-level reports such as the student pre-registration report, course and subject report, student registration report etc. 6. IP-MAP helps users ***to understand the system architecture of the organization***. Through the IP-MAP, they can understand how data is processed and stored in the database and data warehouse. 7. IP-Map helps users ***to visualize the data processing steps to generate the information reports*** using the appropriate data capturing, validating, processing, storing and generating processes. This allows users to identify the critical data, steps, timeline and dependencies for an information report. 8. Users need ***to be assured of the quality of data*** by knowing how complete and accurate the data is before using it for any decision making. A weight is assigned by the users ***to specify the importance or relevance of the data*** when computing the data quality parameters. IP-Map allows the users to do so by modifying the weight of data according to their requirements. 9. IP-Map provides the following two functions of ***time management***: (a) To estimate the time required for generating a set of information reports. (b) To determine the critical path with the timeliness details of each activity through the PERT chart.			

Additional Comments:		
Section C: Business Intelligence Product Map	**Rate**	**Comments**
BIP-Map is built as a multi-layered framework by modifying and integrating the salient modeling and management techniques of BPMN and IP-Map. Multiple BI Dashboards are being attached onto the 1^{st} layer of BIP-Map to provide users the important visual information in a summarized and detailed manner. **BIP-Map Layer 1** is a modified version of the BPMN model while **BIP-Map Layer 2** is a modified version of the IP-Map model. These two layers provide users a clear and proper workflow of the business and information processes that are involved in the Subject Pre-Registration Process respectively. The statements below describe about the features that are available after integrating the three models to construct the BIP-Map. Please **rate the usefulness of the features** towards business decision making.		
10. ***Mapping the business processes (BPMN) with information processes (IP-MAP)*** using the appropriate navigation and highlighting functions: There are many occasions (e.g. failure to meet a certain deadline) where a user needs to identify the detailed data processing steps (Layer 2, IP-MAP) that are related to the business activities (Layer 1, BPMN). 11. ***Bringing the information product (BI Dashboard) from the BIP-MAP Layer 2 (IP-MAP) into Layer 1 (BPMN)*** enables users to view the important information in a visual manner using various kinds of graphs or charts. This is useful for management users to view the summary report of a *business activity*. 12. ***Retrieving and viewing detailed data related to a business process and an information product at the 1^{st} layer of BIP-Map***: The management users can conveniently retrieve very specific data used to generate the information product, when needed.		

13. ***Identifying the (critical) data for each activity***: A weight is automatically calculated for the data based on the frequency and condition of its usage in the BI Dashboard. The weight of data will enable users to identify the importance of each data item. The weight is useful to compute the data quality, i.e., completeness and accuracy of the data. 14. ***Time management*** in BIP-MAP: BIP-MAP allows users to perform time estimation using the PERT chart function. The timing parameters from Layer 2 (IP-Map) can be automatically transformed into Layer 1 (BPMN).		
Additional Comments:		

Section D: Comparison between Models

Please rate the performance of each item in every model with a score (0 – 5) and rank them accordingly (1 – 3). Kindly provide your comments also.

Items to Evaluate		BPMN	IP-Map	BIP-Map
15. *Business Process Management*: To break down the processes into sub tasks by having them to be represented by different modules so that all the processes can be managed easily.	Score			
	Rank			
16. *Business Process Management*: To identify the roles of participants and their interactions with one another for different tasks.	Score			
	Rank			
17. *Business Intelligence Reporting*: To identify the intermediate and final deliverables at different phases of a process.	Score			
	Rank			
18. *Time Management*: To estimate the time that is required to complete each activity available in a process.	Score			
	Rank			

19. *Information Process Management*: To model the processes that describe how data is being captured, validated, processed, stored and utilized to generate the information reports.	Score			
	Rank			
20. *Data Quality Management*: To help users identify the importance of data by providing a weight for each data item.	Score			
	Rank			

Additional Comments:

Section E: Overall Feedback

21. Does the BI Dashboard serve as a useful tool for decision making when the users are able to view the information of an organization in a summarized and detailed manner? Please provide your comments.

    ____

    ____

22. Does the first layer of BIP-Map provide users a clear picture about the business processes that are involved in producing reports used for making decision in an organization? Please provide your comments.

    ____

    ____

23. Does the second layer of BIP-Map provide users a clear picture about how the data is being captured, processed, stored and utilized throughout an organization? Please provide your comments.

    ____

    ____

24. How useful is the integration of the two models and the BI Dashboard in assisting users to perform the decision making process?

    ____

    ____

25. Which areas of the BIP-Map need to be enhanced in order to produce a better framework for Business Intelligence? How should it be improved?

    ____

    ____

Thank you for spending your precious time in doing this survey.

Appendix B: Survey Questions for Online Job Recruitment Company

UNIVERSITI TUNKU ABDUL RAHMAN
Faculty of Information & Communication Technology
Department of Information Systems

Survey Questions

A Business Intelligence (BI) system provides users with actionable information to support decision making process. However, often business users still possess limited picture about the entire information manufacturing process underneath a particular piece of BI product. They are not able to trace the information manufacturing chain, let alone the information quality throughout the information manufacturing process. In response to this, an integrated information manufacturing framework for BI systems, so-called **Business Intelligence Product Map** (**BIP-Map**) has been proposed and developed.

Please take a moment to provide your feedback. Your time spent in doing this survey is highly appreciated. Thank you.

Job Position

SD	D	N	A	SA
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Please tick only one answer at any of the following 5 columns.	SD	D	N	A	SA
1. BI Dashboard is useful in providing information to Managers and Business Analysts during the decision making process of an organisation.					
2. Managers and Business Analysts may possess limited picture about the entire information manufacturing process underneath a particular piece of graph or chart produced in the BI Dashboard.					
3. It is good to have a framework that can allow users to identify the business processes that are related to the information provided to them.					
4. It is good to have a framework that can allow users to visualise the manufacturing processes of information in their organisations especially from a quality aspect.					

5. It is good to have a framework that can allow users to understand the real meaning of data by referring to the metadata available in the framework.					
6. Having such a framework would allow users to easily identify the critical phases that may create data quality problems.					
7. Having such a framework would allow users to implement continuous improvement for the BI needs of their organisations.					
8. The BIP-Map is useful since it provides the workflow of business processes so that users can understand the relationship of all participants and processes for a business transaction in order to implement Business Process Management.					
9. The BIP-Map is useful since it provides the workflow of information processes so that users can identify how the data is being captured, processed, stored and utilised in order to produce better information products.					
10. The BIP-Map is useful since it provides metadata so that users can understand the real meaning of the data before using it to make important decision for their organisations.					

1. Do you think the BIP-Map is useful for business decision making when it is being integrated into a Business Intelligence Dashboard?

2. Does the first layer of BIP-Map provide users a clear picture about the business processes that are involved in producing reports used for making decision in an organization?

3. Does the second layer of BIP-Map provide users a clear picture about how the data is being captured, processed, stored and utilized throughout an organization?

4. Does the third layer of BIP-Map provide users important metadata that can describe the meaning of data available in the database and data warehouse of an organization?

5. Which areas of the BIP-Map need to be enhanced in order to produce a better framework for Business Intelligence?

Appendix C: BIP-MAP for Job Advertisement

This process models the steps that are involved for a job advertisement to be added, updated or deleted from the online job recruitment firm.

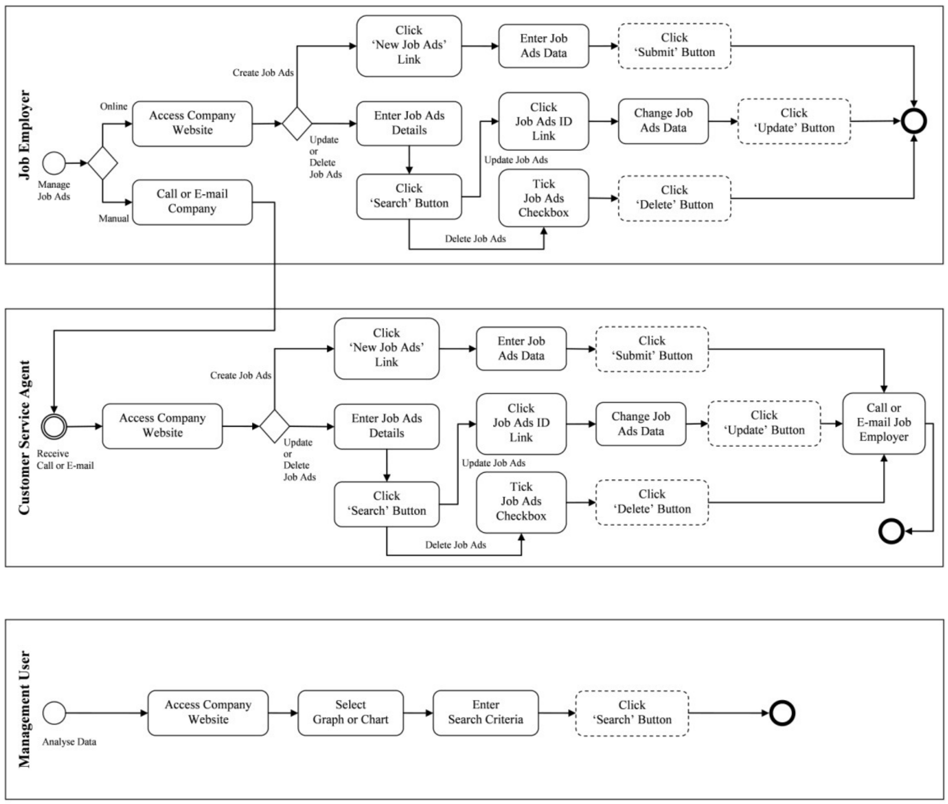


Figure 1: BIP-MAP Layer 1 (Job Advertisement)

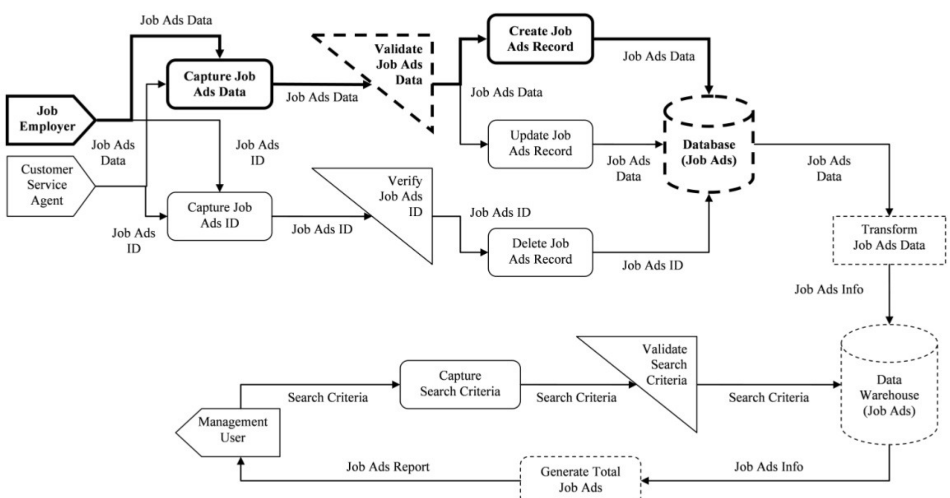


Figure 2: BIP-MAP Layer 2 (Job Advertisement)

The job advertisements available in the online job recruitment firm are compared to the job advertisements available in other media printing companies.

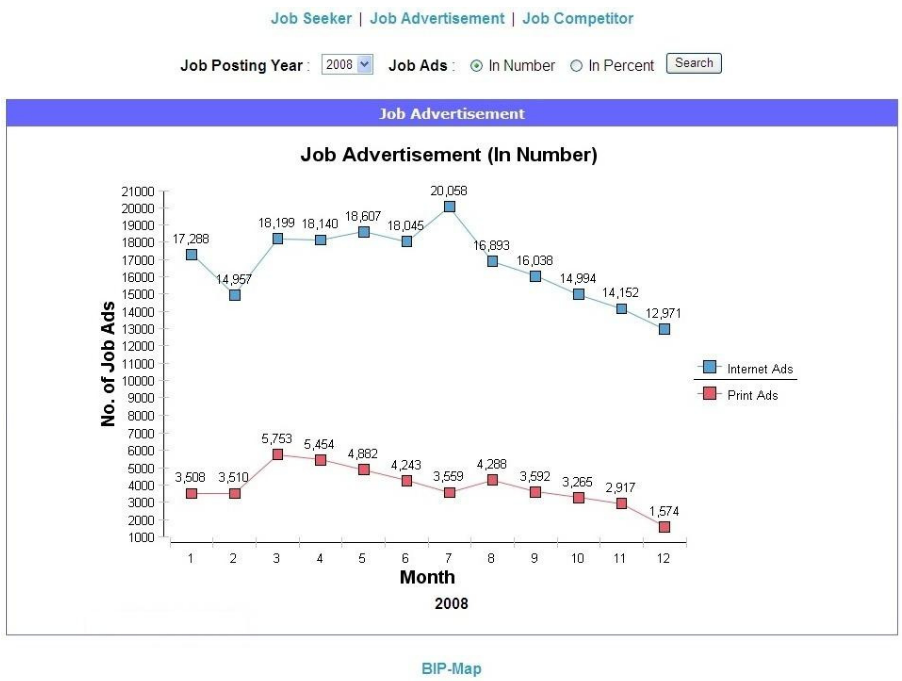


Figure 3: BI Dashboard (Job Advertisement – In Number)

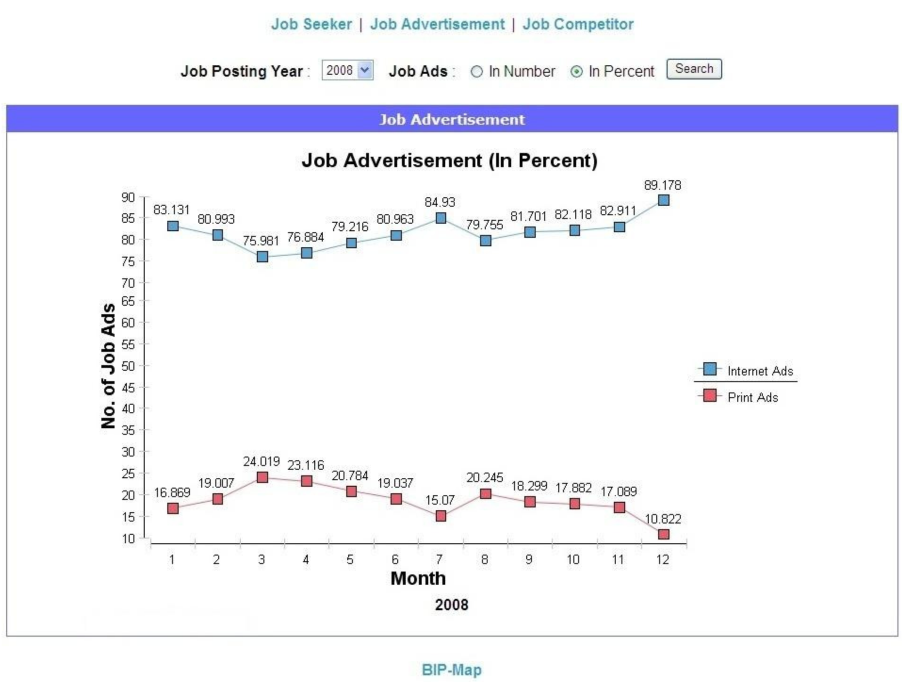


Figure 4: BI Dashboard (Job Advertisement – In Percent)

This metadata describe how the job advertisement data is being validated before storing into the database.

Metadata for Data Validation (Job Advertisement)

Data	Description	Condition	Compulsory	Option
Job Position	– A name that describes the job.	– Not more than 25 alphanumeric characters.	Yes	-
Country	– A country for the job applicant to work.	– Select one option only.		– A list of countries.
State	– A state for the job applicant to work.			– A list of states for the selected country.
City	– A city for the job applicant to work.			– A list of cities for the selected state.
Job Responsibilities	– Information that describes what the job applicant needs to do.	– Not more than 500 alphanumeric characters.	No	-
Job Requirements	– Academic qualification and working experience that the job applicant needs to fulfill.			-
Application Deadline	– The job is closed for application on this date.	– A date that is greater than the current date. – The default selected date is 1 month later than the current date.	Yes	– A javascript calendar with date selection.
Advertisement Status	– Indicates whether the job advertisement is posted online or not. – Job advertisement that is saved but not posted online will not be displayed.	– Select one option only. – Default Value : Post Online		– Post Online – On Hold

Figure 5: Metadata for Data Validation (Job Advertisement)

This metadata describe how the job advertisement data is being transformed from the database into the data warehouse.

Metadata for Data Transformation (Job Advertisement)

Database

advertisement Table	
advertisement_id	int
advertiser_id	int
original_posting_date	datetime
advertisement_status_code	int

advertiser Table	
advertiser_id	int
industry_code	int

ref_industry Table	
industry_code	int
industry_name	varchar(255)

ETL Process

1. Create the **DimIndustry** dimension table.
   - DimIndustry.IndustryKey: Auto-generated
   - ref_industry.industry_code → DimIndustry.IndustryCode
   - ref_industry.industry_name → DimIndustry.IndustryName
2. Create the **DimAdvertiser** dimension table.
   - DimAdvertiser.AdvertiserKey: Auto-generated
   - advertiser.advertiser_id → DimAdvertiser.AdvertiserID
3. Create the **DimPostingDate** dimension table and **FactJobAdvertisement** fact table.
   - DimPostingDate.PostingDateKey: Auto-generated
   - advertisement.original_posting_date → DimPostingDate.PostingDate
   - YEAR(advertisement.original_posting_date) → DimPostingDate.PostingYear
   - DATENAME(MONTH, advertisement.original_posting_date) → DimPostingDate.PostingMonth
   - DAY(advertisement.original_posting_date) → DimPostingDate.PostingDay
   - DimAdvertiser.AdvertiserKey → FactJobAdvertisement. AdvertiserKey
   - DimIndustry.IndustryKey → FactJobAdvertisement.IndustryKey
   - DimPostingDate.PostingDateKey → FactJobAdvertisement. PostingDateKey
   - advertisement.advertisement_id → FactJobAdvertisement.AdvertisementID
   - advertisement.advertisement_status_code → FactJobAdvertisement. AdvertisementStatusCode

Data Warehouse

FactJobAdvertisement Table	
AdvertiserKey	int
IndustryKey	int
PostingDateKey	int
AdvertisementID	int
AdvertisementStatusCode	int

DimAdvertiser Table	
AdvertiserKey	int
AdvertiserID	int

DimIndustry Table	
IndustryKey	int
IndustryCode	int
IndustryName	varchar(255)

DimPostingDate Table	
PostingDateKey	int
PostingDate	datetime
PostingYear	int
PostingMonth	varchar(50)
PostingDay	int

Figure 6: Metadata for Data Transformation (Job Advertisement)

This metadata describe how the job advertisement data is being generated from the data warehouse.

Metadata for Data Generation (Job Advertisement)

Data Warehouse

FactJobAdvertisement Table		
AdvertiserKey	int	-
IndustryKey	int	-
PostingDateKey	int	-
AdvertisementID	int	-
AdvertisementStatusCode	int	1 : Posted Jobs 2 : Not Posted Jobs

DimAdvertiser Table		
AdvertiserKey	int	-
AdvertiserID	int	-

DimIndustry Table		
IndustryKey	int	-
IndustryCode	int	-
IndustryName	varchar(255)	-

DimPostingDate Table		
PostingDateKey	int	-
PostingDate	datetime	-
PostingYear	int	-
PostingMonth	varchar(50)	-
PostingDay	int	-

Report Generation Process

Report Objective :

1. To calculate the number of job advertisements which have been posted by the print and online advertisers.
   - Number of Print Advertisements
   - Number of Online Advertisements
2. To calculate the percentage of job advertisements which have been posted by the print and online advertisers.
   - Percentage of Print Advertisements = Print Ads / Total Ads X 100%
   - Percentage of Online Advertisements = Online Ads / Total Ads X 100%

Default Data Selection Rule :

1. Year(***PostingDate***) = Job Posting Year
2. Jobs Posted Online (***AdvertisementStatusCode*** = 1)
3. Jobs Posted for all Industry Categories

Optional Data Selection Rule :

1. Year(***PostingDate***) = Job Posting Year
2. Jobs On Hold (***AdvertisementStatusCode*** != 1)
3. Jobs Posted for an Industry Category

Report Structure

Chart Title	Job Advertisement (In Number)
Chart Type	Line Graph
Dimension	2D
X-Axis	PostingDate (Month)
Y-Axis	Number of Job Advertisement
Chart Title	Job Advertisement (In Percentage)
Chart Type	Line Graph
Dimension	2D
X-Axis	PostingDate (Month)
Y-Axis	Percentage of Job Advertisement

Figure 7: Metadata for Data Generation (Job Advertisement)

This metadata describe the meaning of each job advertisement data available in the database.

Metadata for Database (Job Advertisement)

advertisement					
Data	**Description**	**Input Type**	**Data Source**	**Data Type**	**Data Code**
advertisement_id	– An auto-generated number used to uniquely identify a job advertisement.	System Input	-	int	-
job_position	– A name that describes the job.	Form Input	Job Position	varchar(50)	-
job_country	– A country for the job applicant to work.		Country	int	– Refer to Country List
job_state	– A state for the job applicant to work.		State		– Refer to State List
job_city	– A city for the job applicant to work.		City		– Refer to City List
job_responsibilities	– Information that describes what the job applicant needs to do.		Job Responsibilities	varchar(250)	-
job_requirements	– Academic qualification and working experience that the job applicant needs to fulfill.		Job Requirements		-
application_deadline	– The job is closed for application on this date.		Application Deadline	datetime	-
advertiser_id	– A company or person who posted the job advertisement. – Automatically set by the system based on the user login ID.	System Input	-	int	– Refer to Advertiser List
original_posting_date	– A date that the job advertisement is posted online. – Automatically set by the system based on the system date.		-	datetime	-
advertisement_status_code	– Indicates whether the job advertisement is posted online or not. – Job advertisement that is saved but not posted online will not be displayed.		Advertisement Status	int	1 : Post Online 2 : On Hold

ref_industry		
Data	**Description**	**Data Type**
industry_code	– An auto-generated number used to uniquely identify an industry.	int
industry_name	– A name that describes the industry.	varchar(255)

ref_advertiser		
Data	**Description**	**Data Type**
advertiser_id	– An auto-generated number used to uniquely identify an advertiser.	int
industry_code	– As a foreign key to identify the industry of an advertiser.	

Figure 8: Metadata for Database (Job Advertisement)

This metadata describe the meaning of each job advertisement data available in the data warehouse.

Metadata for Data Warehouse (Job Advertisement)

FactJobAdvertisement					
Data	**Description**	**Input Type**	**Data Source**	**Data Type**	**Data Code**
AdvertiserKey	– A number used to uniquely identify an advertiser.	Extraction	– DimAdvertiser * AdvertiserKey	int	-
IndustryKey	– A number used to uniquely identify an industry.		– DimIndustry * IndustryKey		-
PostingDateKey	– A number used to uniquely identify a job posting date.		– DimPostingDate * PostingDateKey		-
AdvertisementID	– A number used to uniquely identify a job advertisement.		– advertisement * advertisement_id		-
AdvertisementStatusCode	– Indicates whether the job advertisement is posted online or not. – Job advertisement that is saved but not posted online will not be displayed.		– advertisement * advertisement_status_code		1 : Posted Jobs 2 : Not Posted Jobs

DimIndustry					
Data	**Description**	**Input Type**	**Data Source**	**Data Type**	**Data Code**
IndustryKey	– An auto-generated number used to uniquely identify an industry.	Generation	– Auto-Generated	int	-
IndustryCode	– The original number used to uniquely identify an industry.	Extraction	– ref_industry * industry_code		-
IndustryName	– A name that describes the industry.		– ref_industry * industry_name	varchar(255)	-

Figure 9: Metadata for Data Warehouse (Job Advertisement)

Appendix D: BIP-MAP for Competitor Advertisement

This process models the steps that are involved for a job advertisement belonging to a competitor of the company to be added, updated or deleted from the online job recruitment firm for the purpose of data analysis.

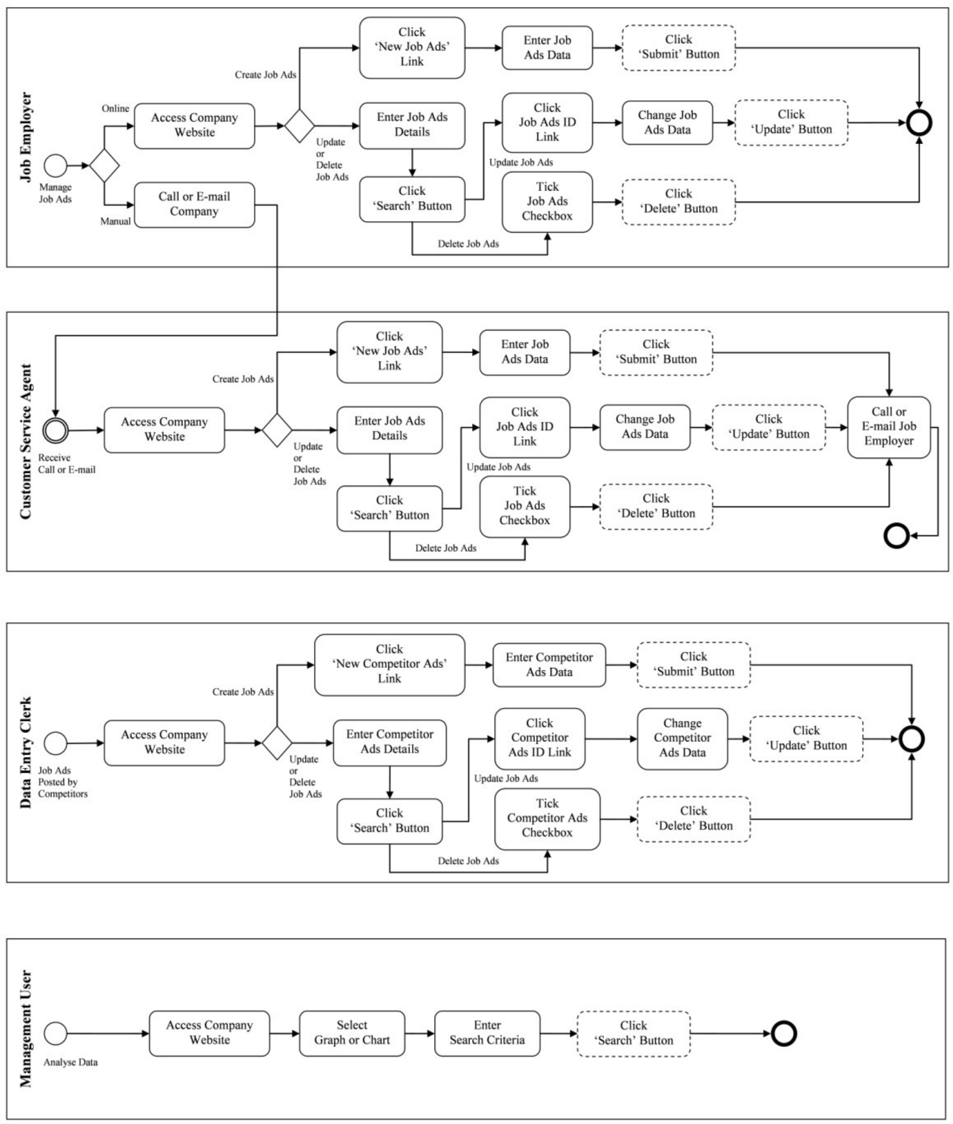


Figure 10: BIP-MAP Layer 1 (Competitor Advertisement)

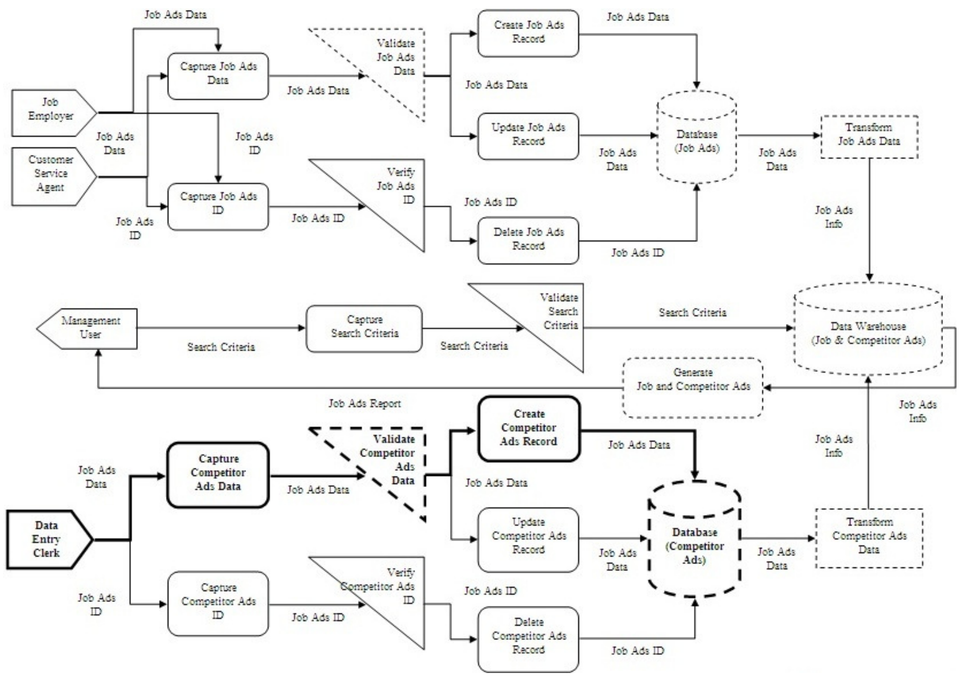


Figure 11: BIP-MAP Layer 2 (Competitor Advertisement)

The job advertisements available in the online job recruitment firm are compared to the job advertisements available in the competitor companies.

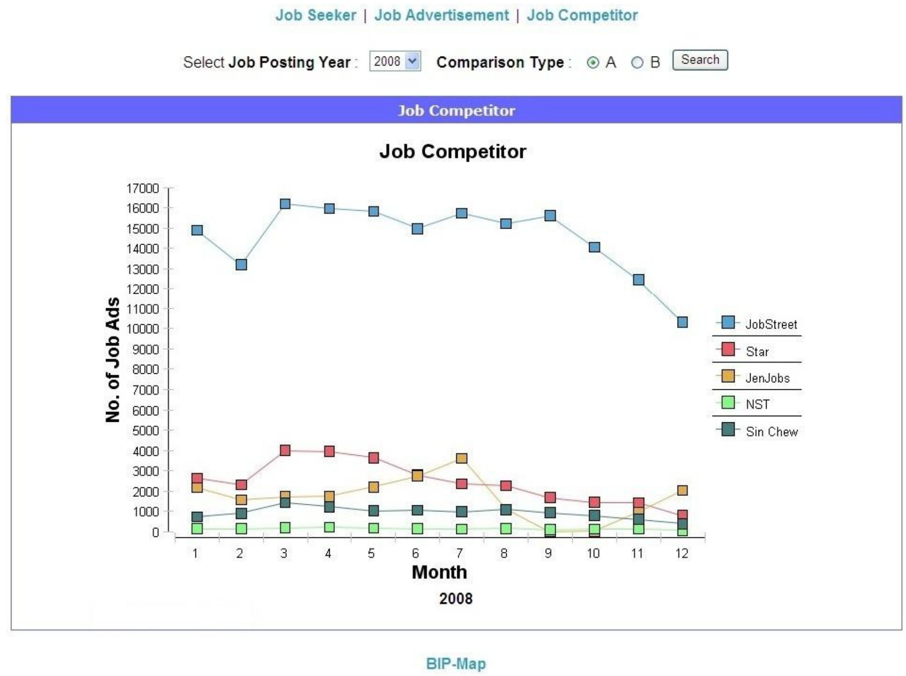


Figure 12: BI Dashboard I (Competitor Advertisement)

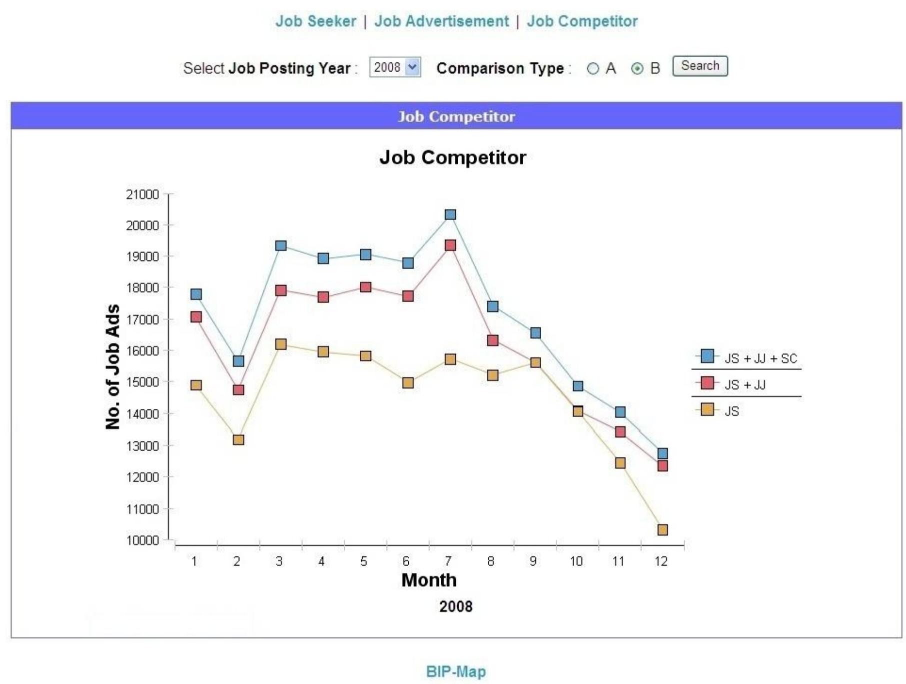


Figure 13: BI Dashboard II (Competitor Advertisement)

This metadata describe how the job advertisement data of the competitor company is being validated before storing into the database.

Metadata for Data Validation (Competitor Advertisement)

Data	Description	Condition	Compulsory	Option
Competitor	– A company that is advertising the job.	– Select one option only.	Yes	– A list of job advertising companies.
Employer	– A company that is offering the job.			– A list of companies.
Job Position	– A name that describes the job.	– Not more than 25 alphanumeric characters.		-
Country	– A country for the job applicant to work.	– Select one option only.		– A list of countries.
State	– A state for the job applicant to work.			– A list of states for the selected country.
City	– A city for the job applicant to work.			– A list of cities for the selected state.
Job Responsibilities	– Information that describes what the job applicant needs to do.	– Not more than 500 alphanumeric characters.	No	-
Job Requirements	– Academic qualification and working experience that the job applicant needs to fulfill.			-
Application Deadline	– The job is closed for application on this date.	– A date that is later than the current date.	Yes	– A javascript calendar with date selection.
Job Posting Date	– A date that the job advertisement is posted online.	– A date that is earlier than the current date.		– A javascript calendar with date selection.

Figure 14: Metadata for Data Validation (Competitor Advertisement)

This metadata describe how the job advertisement data of the competitor company is being transformed from the database into the data warehouse.

Metadata for Data Transformation (Competitor Advertisement)

Database

advertisement Table	
advertisement_id	int
advertiser_id	int
original_posting_date	datetime
advertisement_status_code	int
advertiser Table	
advertiser_id	int
industry_code	int
ref_industry Table	
industry_code	int
industry_name	varchar (255)

ETL Process

1. Create the **DimCompetitor** dimension table.
   - DimCompetitor. CompetitorKey: Auto-generated
   - ref_competitor. competitor_id → DimCompetitor.CompetitorID
2. Create the **DimCompetitorPostingDate** dimension table and **FactCompetitorAdvertisement** fact table.
   - DimCompetitorPostingDate.CompetitorPostingDateKey: Auto-generated
   - competitor_advertisement.posting_date → DimCompetitorPostingDate.CompetitorPostingDate
   - YEAR(competitor_advertisement.posting_date) → DimCompetitorPostingDate.CompetitorPostingYear
   - DATENAME(MONTH, competitor_advertisement.posting_date) → DimCompetitorPostingDate.CompetitorPostingMonth
   - DAY(competitor_advertisement.posting_date) → DimCompetitorPostingDate.CompetitorPostingDay
   - DimCompetitor. CompetitorKey → FactCompetitorAdvertisement.CompetitorKey
   - DimCompetitorPostingDate.CompetitorPostingDateKey → FactCompetitorAdvertisement.CompetitorPostingDateKey
   - competitor_advertisement.competitor_advertisement_id → FactCompetitorAdvertisement.CompetitorAdvertisementID

Data Warehouse

FactJobAdvertisement Table	
AdvertiserKey	int
IndustryKey	int
PostingDateKey	int
AdvertisementID	int
AdvertisementStatusCode	int
DimAdvertiser Table	
AdvertiserKey	int
AdvertiserID	int
DimIndustry Table	
IndustryKey	int
IndustryCode	int
IndustryName	varchar(255)
DimPostingDate Table	
PostingDateKey	int
PostingDate	datetime
PostingYear	int
PostingMonth	varchar(50)
PostingDay	int

Figure 15: Metadata for Data Transformation (Competitor Advertisement)

This metadata describe how the job advertisement data of the competitor company is being generated from the data warehouse.

Metadata for Data Generation (Competitor Advertisement)

Data Warehouse

FactJobAdvertisement Table		
AdvertiserKey	int	-
IndustryKey	int	-
PostingDateKey	int	-
AdvertisementID	int	-
AdvertisementStatusCode	int	1 : Posted Jobs 2 : Not Posted Jobs

FactCompetitorAdvertisement Table		
CompetitorKey	int	-
CompetitorPostingDateKey	int	-
CompetitorAdvertisementID	int	-

DimAdvertiser Table		
AdvertiserKey	int	-
AdvertiserID	int	-

DimCompetitor Table		
CompetitorKey	int	-
CompetitorID	int	-
CompetitorName	varchar(255)	-
CompetitorType	int	1 : Print Advertiser 2 : Online Advertiser

DimIndustry Table		
IndustryKey	int	-
IndustryCode	int	-
IndustryName	varchar(255)	-

DimPostingDate Table		
PostingDateKey	int	-
PostingDate	datetime	-
PostingYear	int	-
PostingMonth	varchar(50)	-
PostingDay	int	-

DimCompetitorPostingDate Table		
CompetitorPostingDateKey	int	-
CompetitorPostingDate	datetime	-
CompetitorPostingYear	int	-
CompetitorPostingMonth	varchar(50)	-
CompetitorPostingDay	int	-

Report Generation Process
Report Objective : 1. To calculate and compare the number of job advertisements which have been posted by the advertisers of JobStreet and its competitors. Default Data Selection Rule : 1. Year(*PostingDate @ CompetitorPostingDate*) = Job Posting Year 2. Jobs Posted Online (*AdvertisementStatusCode* = 1) Optional Data Selection Rule : 1. Year(*PostingDate @ CompetitorPostingDate*) = Job Posting Year 2. Jobs Posted by different type of Competitors (Print Advertisers or Online Advertisers)

Report Struture	
Chart Title	JobStreet vs. Competitor
Chart Type	Multi Lines Graph
Dimension	2D
X-Axis	PostingDate @ CompetitorPostingDate (Month)
Y-Axis	Number of Job Advertisement

Figure 16: Metadata for Data Generation (Competitor Advertisement)

This metadata describe the meaning of each competitor's job advertisement data available in the database.

Metadata for Database (Competitor Advertisement)

competitor_advertisement					
Data	**Description**	**Input Type**	**Data Source**	**Data Type**	**Data Code**
competitor_advertisement_id	– An auto-generated number used to uniquely identify a job advertisement.	System Input	-	int	-
c_job_position	– A name that describes the job.	Form Input	Job Position	varchar(50)	-
c_job_country	– A country for the job applicant to work.		Country	int	– Refer to Country List
c_job_state	– A state for the job applicant to work.		State		– Refer to State List
c_job_city	– A city for the job applicant to work.		City		– Refer to City List
c_job_responsibilities	– Information that describes what the job applicant needs to do.		Job Responsibilities	varchar(250)	-
c_job_requirements	– Academic qualification and working experience that the job applicant needs to fulfill.		Job Requirements	varchar(250)	-
c_application _deadline	– The job is closed for application on this date.		Application Deadline	datetime	-
competitor_id	– A company or person who posted the job advertisement.		Competitor	int	– Refer to Competitor List
employer_id	– A company that is offering the job.		Employer		– Refer to Employer List
posting_date	– A date that the job advertisement is posted online.		Job Posting Date	datetime	-

ref_competitor		
Data	**Description**	**Data Type**
competitor_id	– An auto-generated number used to uniquely identify a competitor.	int
competitor_name	– The name of a competitor.	varchar(50)

Figure 17: Metadata for Database (Competitor Advertisement)

This metadata describe the meaning of each competitor's job advertisement data available in the data warehouse.

Metadata for Data Warehouse (Competitor Advertisement)

FactJobAdvertisement					
Data	**Description**	**Input Type**	**Data Source**	**Data Type**	**Data Code**
AdvertiserKey	– A number used to uniquely identify an advertiser.	Extraction	– DimAdvertiser * AdvertiserKey	int	-
IndustryKey	– A number used to uniquely identify an industry.		– DimIndustry * IndustryKey		-
PostingDateKey	– A number used to uniquely identify a job posting date.		– DimPostingDate * PostingDateKey		-
AdvertisementID	– A number used to uniquely identify a job advertisement.		– advertisement * advertisement_id		-
AdvertisementStatusCode	– Indicates whether the job advertisement is posted online or not. – Job advertisement that is saved but not posted online will not be displayed.		– advertisement * advertisement_status_code		1 : Posted Jobs 2 : Not Posted Jobs

DimIndustry					
Data	**Description**	**Input Type**	**Data Source**	**Data Type**	**Data Code**
IndustryKey	– An auto-generated number used to uniquely identify an industry.	Generation	– Auto-Generated	int	-
IndustryCode	– The original number used to uniquely identify an industry.	Extraction	– ref_industry * industry_code		-
IndustryName	– A name that describes the industry.		– ref_industry * industry_name	varchar(255)	-

Figure 18: Metadata for Data Warehouse (Competitor Advertisement)