# Improving Business Intelligence Traceability and Accountability: An Integrated Framework of BI Product and Metacontent Map

*Chin-Hoong Chee, Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman, Kampar, Perak, Malaysia.*

*William Yeoh, School of Information and Business Analytics, Deakin University, Burwood, VIC, Australia*

*Shijia Gao, Faculty of Information Technology, Monash University, Caulfield East, VIC, Australia*

*Gregory Richards, Telfer School of Management, University of Ottawa, Ottawa, Ontario, Canada*

## ABSTRACT

*A Business Intelligence (BI) system provides users with multi-dimensional information (a so-called 'BI product') to support decision-making. However, existing BI systems overlook the lineage metadata which supports individual data quality dimensions such as data believability and ease of understanding. Using a design science research paradigm, this paper proposes and develops an integrated framework (known as BI Product and Metacontent Map - 'BIP-Map') to facilitate the traceability and accountability of BI products. Specifically, the business workflow layer of the integrated framework is modelled using business process modelling notation, and an information product map is used to model the second layer's information manufacturing process, whilst the third layer represents the metacontent detail of the data validation stage, from source system through to ETL, to the data warehousing stage. Also, the authors develop a BIP-Map informed prototype in collaboration with an online job advertising firm, the framework then being validated by key BI stakeholders of the firm. The integrated framework addresses individual-related data quality issues and builds user confidence by enhancing the traceability and accountability of a BI product.*

*Keywords:* *Business Intelligence, Business Workflow, Design Science Metadata, Information Manufacturing, Metadata*

DOI: 10.4018/jdm.2014070102

## INTRODUCTION AND RESEARCH MOTIVATION

In recent years, business intelligence (BI) tools have emerged as one of the top spending priorities for many CIOs (Gartner 2013a). It is reported that worldwide business intelligence software revenue surpassed US$12 billion dollars in 2012 (Gartner 2013b). Organisations are willing to invest heavily in BI because this technology enables companies to make fact-based decisions which can increase revenue, cut costs, and provide competitive advantage (Bogza & Zaharie 2008). According to Wixom and Watson (2010, p.14), business intelligence is "a broad category of technologies, applications, and processes used for gathering, storing, accessing, and analysing data to help its users make better decisions". In other words, BI is a system that allows users to leverage disparate data sources so that they can make informed business decisions (Elbashir et al., 2008; Negash 2004). To do so, BI systems are designed to provide decision-makers with BI products[1] delivered at the right time, at the right place, and in the correct form (Negash 2004). A BI product is akin to manufacturing a physical product in the factory because most of the processes used in product quality-control are applicable in producing higher quality information (Ballou et al., 1998; Shankaranarayanan et al., 2003).

While the BI market appears vibrant and while the importance of BI systems is more widely accepted, a weakness of current systems is that data are treated too technically without considering the 'soft' aspects of the information; for instance, aspects such as data lineage and reliability are often ignored (Foshay 2005; Foshay et al., 2007; Shankaranarayanan et al., 2003). A data-driven BI system that concentrates mainly on data itself will yield lower levels of efficiency in business decision-making (Ou & Peng 2006). This is because business decision-making is a complex process which involves not only hard data, such as sales volumes, but also other considerations such as the information manufacturing process and standard operating procedures (SOP) of the company. For example, any changes in the SOP or business workflow may affect the operations of the business which, in turn, affects how data are generated; indeed, such changes can affect the reliability of the BI product manufacturing process itself. Moreover, contemporary managers are forced to make optimised decisions in the shortest possible time in response to today's highly competitive and fast-paced business environment (Delen & Pratt 2006). Yet, business users still rely on IT to supply the metadata/metacontent[2], a time-consuming process which can delay them from acting/reacting promptly. A problem with current BI systems is that business users typically engage only at the complex analysis layer without access to adequate metacontent to support their decision-making. Dynamic decision-support demands adequate lineage metadata[3] as well as sufficient knowledge of business workflow and BI product manufacturing processes.

Due to inadequate lineage metadata in existing BI systems (Foshay 2005; Foshay et al., 2007), questions have been raised about the ability of business users who are not technically-savvy to understand the BI environment or to adequately comprehend the highly-granular information provided by BI technologies. Consequently, the problems associated with individual-related data-quality dimensions, such as ease of understanding, data believability, and accessibility, are becoming more pronounced (Wang & Strong 1996). Similarly, there are problems in understanding the relationships of all participants and processes for a business transaction (Tan et al., 2008). Therefore, the provision of comprehensive metadata to enhance the understanding of BI products has grown in importance (Foshay 2005; Foshay et al., 2007; Shankaranarayanan & Even 2006).

While 'metadata' has been defined as 'data about the data', this paper uses the definition suggested by Dempsey and Heery (1998, p. 155) who describe it as "data that is associated with objects which relieves their potential users of having full advance knowledge of their existence or characteristics." This definition addresses individual-related data quality challenges, especially BI product believability, ease of understanding, and metadata accessibility, all of which are paramount for heightening BI users' acceptance (Wang & Strong 1996; Foshay 2005; Foshay et al., 2007). As the quantity and complexity of corporate data has increased dramatically, and as data becomes more widely distributed throughout organisations, users at all management levels are required to make decisions in more responsive and dynamic ways (Shankaranarayanan et al., 2003). However, the perceived quality of data is affected by the way it is applied by users because the same data may be viewed in different perspectives by different people in different situations (Shankaranarayanan & Cai 2006). If users of data interpret the information incorrectly then there may be severe consequences to the business of an organisation. Thus, it is imperative that metadata should be implemented properly at various levels in a BI system in order to ensure that data are understood and used correctly when decisions are made.

Further, data governance and stewardship have been accorded particular prominence in many organisations recently because the quantity of corporate data continues to expand exponentially (Tallon 2013). In addition, regulatory mandates - such as the Sarbanes-Oxley (SOX) Act and Basel II - have increased the attention now being paid to data governance (Khatri & Brown 2010). However, a recent survey has indicated that most of the resources allocated by organisations to data governance provide little value (Aiken et al., 2011). To achieve better data governance it is important to help end-users access metacontent in order to make informed decisions that are also in compliance with regulatory requirements. However, while much literature has been published on the subjects of data quality and governance, little attention has been paid to the provision of comprehensive lineage metadata for the entire BI product manufacturing process. Although useful, the Verifiable Intelligence Products (VIP) framework developed by Zhu and Wang (2010) is targeted to the intelligence community, such as the Central Intelligence Agency of the USA, for the purpose of tracing the sources and production processes of an intelligence product. Their data emanates from a variety of external sources and are characterised by large volumes of semi-structured and unstructured data, so-called 'big data'. Their VIP framework does not take into account the internal, structured data common to an enterprise's BI systems. Drawing on the IP-Map of Shankaranarayan et al. (2003), Shankaranarayanan and Cai (2006) included a comprehensive metadata repository within the IP-Map framework to assist in computing data quality dimensions for implementation of total data quality management capabilities. While the metadata repository is helpful in communicating data quality metadata, it does not consider the specific metadata components for data validation, extraction, transformation, loading and storage in both the database and data warehouse of an organisation – component that are critical for better understanding of BI products.

Existing research has proposed that the management of both business and IT processes need to be integrated with BI in order to better manage business performance (Shi & Lu 2010). Therefore, in this paper we present an integrated framework of BI Product and Metacontent Map ('BIP-Map') which serves as a mechanism to provide the context for the BI product. Specifically, through a design science research approach (Hevner & Chatterjee 2010) we integrate the modelling techniques used in business process modelling, the information product map (IP-Map) of total data quality management, and the metacontent representation so as to construct a three-layered integrated framework. Moreover, we demonstrate and validate the BIP-Map framework in the real-world setting of an online job advertising firm.

It is expected that this research will advance BI lineage research and, more broadly, that it will contribute to the field of data quality and governance. This research builds on and extends the foundational work of Shankaranarayan et al. (2003) by supplementing their IP-Map concept with business workflow and metadata associated with BI products. Shankaranarayan et al. (2003) focused on data-quality management but they did not consider other important aspects in data governance, such as business processes and metadata components. Thus, this research addresses these particular aspects of data governance which have been neglected in prior studies. Specifically, the integrated BIP-Map framework will assist BI stakeholders to make timely and informed decisions because the framework will enable users to better visualise how their business and information manufacturing processes are interrelated. In addition, the framework provides comprehensive metacontent that will allow users to better understand the business rules and metadata of the BI products from data validation through to the data warehousing stage before making any important decisions. Furthermore, the framework enables BI stakeholders to identify the critical parts of a process or system that may create the problem of poor data quality, and by doing so the business intelligence of an organisation can be improved continually.

The remainder of this paper is organised as follows. The next section provides the research paradigm and introduces the integrated BIP-Map framework and the relevant literature. In the following section we demonstrate the utility of the framework within the setting of an online job advertising firm, followed by interview findings with the key BI stakeholders of the firm. The final section provides a summary of the study; it reflects on the design science paradigm and highlights the research contributions.

## RESEARCH PARADIGM

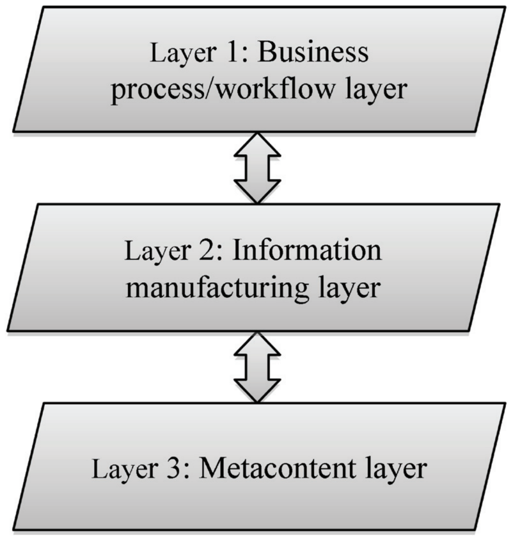
This study applies a design science research paradigm. According to Hevner et al. (2004), the objective of design science is to create artefacts (broadly defined, including constructs, models, methods, and instantiations) that are useful. This research demonstrates the applicability and utility of an integrated framework of BIP-Map (which is classified as a model) and a BIP-Map informed prototype system (i.e., an instantiation). The purpose of the design and development of these two artefacts is to enable the users to better understand the BI product and its manufacturing process and hence improve the efficiency and effectiveness of their decision-making. The design science paradigm aligns with this research's purpose perfectly given that the focus is to design and test artefacts that improve BI system understanding and use. Hevner et al.'s (2004) study specified a set of guidelines for the conduct and assessment of information systems (IS) design-science research that has been widely accepted. The detailed assessment of this study against Hevner et al.'s guidelines is provided in the final section of this paper.

## INTEGRATED FRAMEWORK OF BIP-MAP

Based on business process modelling notation, information product map and metacontent, we develop the integrated framework known as a Business Intelligence Product and Metacontent Map (BIP-Map). The purpose of this integration is to produce a comprehensive framework which is able to enhance the traceability and accountability of BI products so that the end-users can make more informed and timely decisions for their business. As depicted in Figure 1 below, the BIP-Map is built as a multi-layered framework where the business processes/workflow, information modelling processes, and metacontent is integrated into three different layers, with business workflow as the first layer, information manufacturing as the second layer, and metacontent as the third layer. The following provides the relevant literature and justification of why these three layers are designed in this specific order.

A business process consists of "a set of activities that are performed in coordination in an organisational and technical environment to realise a business goal" (Weske 2007, p.5). It is crucial to identify clearly the activities and participants involved in a business workflow, since a BI product can only be analysed if its business workflow is being modelled

*Figure 1. High-level BIP-Map framework*




correctly (Aguilar-Saven 2003). To represent the processes or workflows of an organisation, business process modelling notation[4] (BPMN) is commonly used so that current business processes or workflow may be analysed and improved in the future. In this research, BPMN, a well-supported standard modelling notation has been applied accordingly. BPMN has been unanimously endorsed as the de facto standard notation by the BPM Initiative Notation Working Group, which represents a large segment of the business workflow modelling community. After reviewing various notations and methodologies, the BPM Initiative Notation Working Group consolidates the best ideas from the divergent notations into a single, standard notation - BPMN. Based on flow-charting techniques, BPMN provides a standardised language for business and technical developers (Wang et al., 2006; White 2004). Thus, BPMN has been integrated into the BIP-Map framework. In other words, the BIP-Map layer one presents the business workflow to the users so that they can understand the relationship of all the participants and processes for a business transaction. Furthermore, when the activities and participants of a specific business workflow are defined at this layer, various manual tasks can be automated and, thus accomplished in a more efficient manner. It also has the added benefit of facilitating rapid changes and improvements to the business workflow. This can be achieved because the interaction of activities between different participants for a business workflow is indicated in this layer.

Although BPMN is a very useful modelling technique for managing business processes as it presents to the users a clear workflow of the transactions involved in the business, it does not, however, describe the manufacturing process of data. Therefore, users will not understand how corporate data are captured, processed, stored and utilised. This problem can be addressed, in part, by the use of an information product map[5] (IP-Map) which describes the manufacturing processes of data in a detailed manner. Originating from the total data quality management concept, the purpose of an IP-Map is to deliver high quality information to users by treating data and information in much the same way as other products used in the manufacturing industry (Wang 1998; Wang et al., 1998; Shankaranarayanan et al., 2000; Shankaranarayanan 2005). It allows people to comprehend, evaluate, and describe how information is compiled. However, a limitation of the IP-Map is that it does not provide descriptions of the business processes that are related to the manufacturing or utilisation of data in an organisation. So, users are still unable to identify the business processes that have an effect on corporate data. Therefore, the adoption of business process modelling and information product map will complement each other. Applying the IP-Map concept, BIP-Map layer two illustrates the information manufacturing processes so that users can identify how the data are being captured, processed, stored and utilised in order to produce better information products. Additionally, the IP-Map enables end-users to visualise the manufacture of an information product (Wang 1998; Wang et al., 1998; Shankaranarayanan et al., 2000; Shankaranarayanan 2005), and to appraise its data quality in order to implement continuous improvement (Fisher et al., 2006).

By providing the content and context of the data directly to users, the quality of the data, especially in terms of ease of understanding, data believability, traceability and accountability is greatly increased. Therefore, the purpose of BIP-Map layer three is to provide metacontent to users so they can understand the business rules and associated metadata of the BI products before using them to make decisions. Traditionally, metacontent is owned and used by technical personnel; but through the provision of a BIP-Map framework, business users are exposed to such metacontent in a timely manner which would improve their BI understanding and decision-making process. The metacontent of layer three is grouped into a few categories according to the entire information manufacturing processes, from data entry, extract, transform and load (ETL) stage, until the data warehousing stage (Ballard et al., 2006; Vaduva and Dittrich 2002). They include metacontent for data filtering/validation, data transformation, database, and for the data warehouse. By referring to the third layer of BIP-Map, BI users can know in detail how data are being validated by the data-filtering process when they are captured from source systems. Knowing how data are validated is vital because it can ensure that only data that are accurate and useful to the system are retained. Moreover, this layer is a useful reference for the technical personnel, such as database administrators and software developers, since metacontent is provided for the transformation of data from a database into a data warehouse. Table 1 below provides a summary of the main components of the BIP-Map framework and the relevant literature which serve as the theoretical foundation of this research.

The advantage of having a hierarchical, integrated framework is that different levels of details are systematically segregated into different layers (from high-level BI products and business processes to the underlying information manufacturing processes). For decision-making, management users can understand the BI products that are supplied by the BI systems in a more efficient manner. Managers are able to drill down into the detailed information manufacturing layer and the associated metacontent layer in order to identify how the data are being captured, validated, processed, stored, transformed and generated throughout the organisation. Apart from this, the business workflow layer helps the technical BI developers/users to understand the business considerations behind certain data, whilst the information manufacturing layer enables the business users to gain insight into the logic behind the information generated for them. Thus, the integrated BIP-Map framework enables users to have a comprehensive understanding of the BI products from the integrated perspectives of business workflows, information manufacturing and associated metacontent.

## BIP-Map Demonstration

To illustrate the utility of the BIP-Map framework, a BIP-Map informed prototype has been developed in collaboration with an online job advertising firm. First, the system requirements for the executive BI Dashboard were gathered through a series of interviews with the key BI stakeholders. Next, the BI Dashboard was developed using PHP programming language and an open source BI tool, referred to as Business

*Table 1. BIP-Map framework and the main components*

| BIP-Map layer | Main component | Representation method | Purpose | References |
|---|---|---|---|---|
| One | Business process/ workflow modelling | Business process modelling notation (BPMN) | Describe the activities and participants of a business workflow relating to the manufacturing of a BI product. | Weske 2007; White 2004; Wang et al., 2006 |
| Two | Information manufacturing modelling | Information product map (IP-Map) | Describe the manufacturing processes related to the information product, so BI users are able to visualise the steps involved in capturing, processing, and storing that BI product. | Wang 1998; Wang et al., 1998; Shankaranarayanan et al., 2000; Shankaranarayanan 2005; Shankaranarayan et al., 2003 |
| Three | Provision of metacontent | Metacontent for data filtering/validation, metacontent for data transformation, metacontent for database, and metacontent for data warehouse | Provide the metacontent information like business rules and metadata so BI users can know in detail how data are being validated by the data-filtering process from source systems to data warehousing stage. | Foshay 2005; Foshay et al., 2007; Ponniah 2001 |

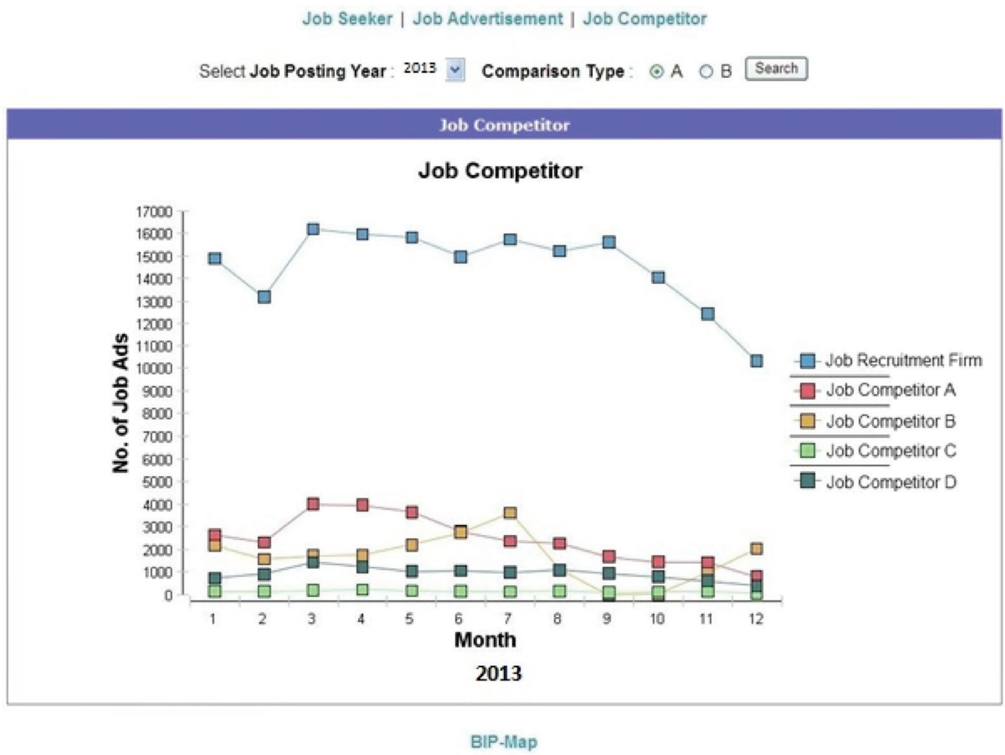
Intelligence and Reporting Tools (BIRT). Both the PHP and the BIRT technologies were used to query the data from the firm's Microsoft SQL Server database into the BI Dashboard. The BIP-Map was attached to the BI Dashboard where the users can click on the link located below any chart available in the BI Dashboard. The following sub-sections illustrate the three-layered BIP-Map framework based on a BI Dashboard of the online job advertising firm.

## *First Layer of BIP-Map: Business Process/Workflow Layer*

Figure 2 above shows an example of a graph available in the BI Dashboard that displays the total number of job advertisements posted by the online job advertising firm and its competitors. This graph enables the BI stakeholders to compare the amount of job advertisements posted by different competitors. It is important for them to identify the number of job advertisements posted each month so that they can execute the appropriate strategies to compete with the competitors.

When users click on the BIP-Map link located below the chart on the BI Dashboard, the first layer of the BIP-Map for that particular information will be invoked. This first layer is based on BPMN and is used to describe the business workflow that is related to the BI product. By referring to the first layer, end users can readily identify the activities and participants of a business workflow relating to the manufacturing of a BI product. Without this layer, a wrong decision could be made because different sets of data in the entire organisation could have similar attribute names and this might mislead users in simply using the data for conducting business decision-making within the wrong context. Hence, understanding the business activities and participants involved is crucial. Figure 3 shows the first layer of a BIP-Map for the job competitor graph. It indicates how the job employers, customer service agents and data entry clerks are involved with the creation of a job advertisement in the firm. When the employer wants to create, update or delete a job advertisement, he can accomplish this by accessing the firm's website or contact-

*Figure 2. Job Competitor's graph from BI dashboard*

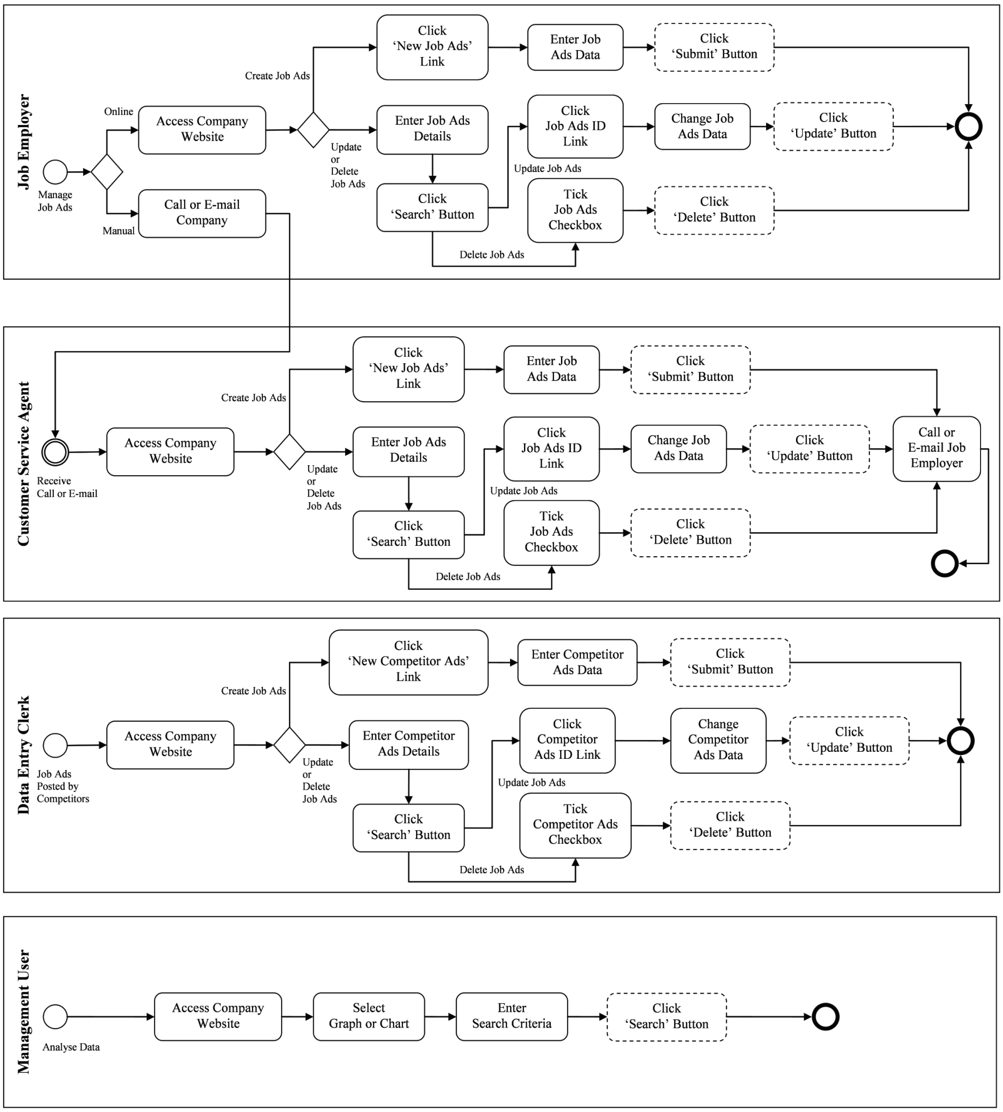



ing the customer service agent through phone or e-mail. The data entry clerks are responsible for entering the job advertisement details of their competitors into the system. As soon as the job advertisement details are available in the system, the management users will access the BI products to support their business analysis and decision-making. The activity represented by a dashed line is expandable, and users can click on it to access the second layer (information manufacturing layer) of the BIP-Map for that particular activity.

## Second Layer of BIP-Map: Information Manufacturing Layer

When users click on any expandable process on the first layer of a BIP-Map, the second layer will then be invoked. The second layer is based on the information product map (IP-Map) of Wang (1998), Wang et al. (1998) and Shankaranarayanan et al. (2000). An IP-Map is used to describe the processes related to the information product, and users are able to visualise the steps involved in capturing, processing, and storing that product (Wang 1998; Wang et al., 1998; Shankaranarayanan et al., 2000; Shankaranarayanan 2005). This helps end-users to make good use of the data in their organisations for business analysis and decision-making. Similarly, it facilitates changes to improve the workflow of information processes when the users are able to identify how the data are being captured, processed, stored and utilised throughout the organisation. Figure 4 below shows the second layer of a BIP-Map for the job competitor graph, which indicates the information manufacturing processes of all the job advertisements posted by the online job advertising firm and its competitors. The job advertisement can be created, updated and deleted by the job employer, customer service agent and data entry clerk. When they create or update a job advertisement, the

*Figure 3. First layer (business workflow) of BIP-Map for Job Competitor graph*




job advertisement data will be captured and validated before it is created or updated in the database. If they delete a job advertisement, the job advertisement ID will be captured and verified before it is deleted from the database. Then, the data will be extracted, transformed and loaded (ETL) from the database into the data warehouse in a multi-dimensional form to support the decision-making process. When the management users view the information for BI reporting purpose, the search criteria will be captured and validated in order to retrieve the relevant data to be included in the BI product. The building block represented by a dashed line is expandable, and users can click on it to access the third layer (metacontent layer) of the BIP-Map for that particular building block.

### *Third Layer of BIP-Map: Metacontent Layer*

Clicking on one of the dashed line building blocks in the second layer of a BIP-Map will invoke a third layer which is based on the concept of BI metacontent representation, and comprises a set of metacontent such as business rules and metadata involved with the manufacturing of the information product from source systems to end BI products. Figures 5 to 8, respectively, represents the third layer of a BIP-Map for the job advertisement graph that shows the metacontent for the data entry/validation stage, metacontent for the ETL stage, metacontent for the storage of data in the database, and metacontent for the data warehouse. Based on this metacontent, managers are able to identify the business rules and criteria used to generate a report from the BI system. Knowing such information will help them more precisely interpret the reports and thus make informed decisions for their organisations. In addition, metacontent on the storage of data in databases and data warehouses are also a useful reference resource for the technical personnel to improve corporate data governance.

## Framework Validation

This research puts forward an integrated BIP-Map framework and demonstrates the utility of the framework within an online job advertising setting in collaboration with JobCo, a medium-sized company based in Southeast Asia that provides services to thousands of corporate customers. A typical service provided by JobCo is online job advertisement posting on its websites. Its corporate customers (i.e. employers) can directly access the firm's online platform and post job advertisements, or alternatively, a customer service agent could post job advertisements on behalf of the employers. The firm also collects job advertising data of its

*Figure 4. Second layer (information manufacturing) of BIP-Map for Job Competitor graph*

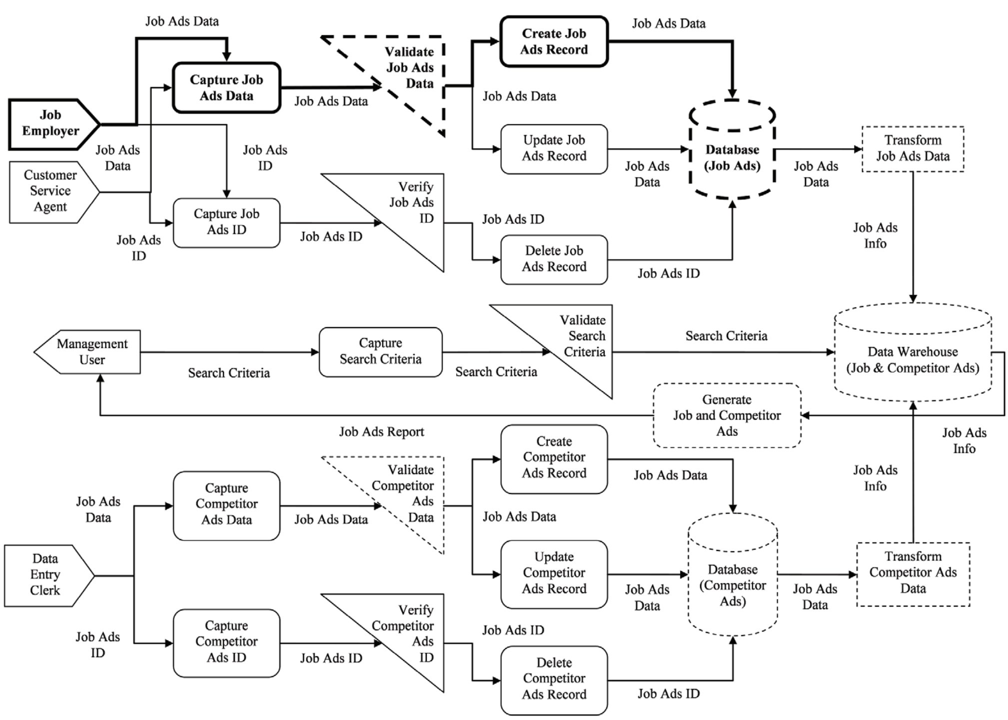


*Figure 5. Third layer of BIP-Map for Job Advertisement graph (data entry/data validation)*

| 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.<br>– 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.<br>– Job advertisement that is saved but not posted online will not be displayed. | – Select one option only.<br>– Default Value : Post Online | | – Post Online<br>– On Hold |

*Figure 6. Third layer of BIP-Map for Job advertisement graph (ETL stage)*

**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 |

main competitors and tracks its monthly sales performance against its competitors using a business intelligence tool.

To validate the usefulness of the BIP-Map, we engaged a panel of BI stakeholders of the online job advertising firm. Specifically, semi-structured interviews were conducted with the firm's seven BI stakeholders, who are also the key BI users of the medium-sized company. They included the regional process improvement manager, the quality and administration manager, the management information systems manager, the chief executive officer, the chief operations officer, the chief information officer, and the chief technology officer. The interview sought to elicit detailed opinions and comments specific to the subject of BIP-Map for the firm's BI systems. Unlike a focus group, an interview method is particularly suitable for this situation because it facilitates personal contact between participants but avoids the possibility of dominance by opinion-leaders, and so ensures democratic participation (Carson et al., 2001). Face-to-face interviews with BI stakeholders, each of which lasted 40 to 50 minutes, were conducted on scheduled times at their respective

*Figure 7. Third layer of BIP-Map for job advertisement graph (Database)*

**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.<br>– 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.<br>– Automatically set by the system based on the system date. | | - | datetime | - |
| ***advertisement_status_code*** | – Indicates whether the job advertisement is posted online or not.<br>– Job advertisement that is saved but not posted online will not be displayed. | | Advertisement Status | int | 1 : Post Online<br>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.Third layer of BIP-Map for Job Advertisement graph (data warehouse)*

**FactJobAdvertisement**

| Data | Description | Input Type | Data Source | Data Type | Data Code |
|---|---|---|---|---|---|
| ***AdvertiserKey*** | – A number used to uniquely identify an advertiser. | Extraction | – DimAdvertiser<br>* AdvertiserKey | int | - |
| ***IndustryKey*** | – A number used to uniquely identify an industry. | | – DimIndustry<br>* IndustryKey | | - |
| ***PostingDateKey*** | – A number used to uniquely identify a job posting date. | | – DimPostingDate<br>* PostingDateKey | | - |
| ***AdvertisementID*** | – A number used to uniquely identify a job advertisement. | | – advertisement<br>* advertisement_id | | - |
| ***AdvertisementStatusCode*** | – Indicates whether the job advertisement is posted online or not.<br>– Job advertisement that is saved but not posted online will not be displayed. | | – advertisement<br>* advertisement_status_code | | 1 : Posted Jobs<br>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<br>* industry_code | | - |
| ***IndustryName*** | – A name that describes the industry. | | – ref_industry<br>* industry_name | varchar(255) | - |

**DimAdvertiser**

| Data | Description | Input Type | Data Source | Data Type | Data Code |
|---|---|---|---|---|---|
| ***AdvertiserKey*** | – An auto-generated number used to uniquely identify an advertiser. | Generation | – Auto-Generated | int | - |
| ***AdvertiserID*** | – The original number used to uniquely identify an advertiser. | Extraction | – advertisement<br>* advertisement_id | | - |

**DimPostingDate**

| Data | Description | Input Type | Data Source | Data Type | Data Code |
|---|---|---|---|---|---|
| ***PostingDateKey*** | – An auto-generated number used to uniquely identify a job posting date. | Generation | – Auto-Generated | int | - |
| ***PostingYear*** | – The year of a job posting date. | Transformation | – advertisement<br>* original_posting_date | | - |
| ***PostingMonth*** | – The month of a job posting date. | | | varchar(50) | - |
| ***PostingDate*** | – The day of a job posting date. | | | int | - |

work sites. During the interview, the researchers demonstrated the working prototype of the BIP-Map and the interviewees were then asked to comment on the artefacts based on a pre-defined list of questions. Hence, it allowed for some latitude in getting detailed information from the interviewees during the interview but, at the same time, general questions could be asked so that all participants were treated in a consistent manner.

The feedback received was instructive and validated the BIP-Map framework. Interview panellists were unanimous in the view that the framework is useful for timely and informed business decision-making when it is integrated into a BI Dashboard. This is because a BIP-Map provides the BI users with lineage metadata – that is, a clear picture of the business workflow and the information manufacturing processes, and also a metacontent understanding on the data tabulated into the various BI reports. Moreover, the panellists recognised that a BIP-Map framework enables users to gain more insight into the displayed reports or charts in a BI Dashboard. One interviewee stated,

*It helps in the understanding of how the information came about, and provides a mechanism for improving the data capturing and information generation processes.*

In other words, having such a three-layered integrated framework would allow BI users to readily identify data quality problems and the personnel responsible for each information product. One panellist pointed out that a BI Dashboard alone is not sufficient for users to fulfil the BI needs of their organisation. Therefore, the BIP-Map framework will serve as a comprehensive framework to address individual-related data quality issues, such as BI product believability, ease of understanding and metadata accessibility, and to provide detailed information that is important for prompt business analysis and timely decision-making. In addition, the panellists asserted that individual end-users may have only a limited knowledge of the information manufacturing process that underlies a particular BI product produced in the Dashboard. One interviewee mentioned that,

*This (i.e. BIP-Map) provides the technical team with some form of documentation of the report structure which can be controlled in order to ensure the report's integrity. It is also useful when the technical personnel, unfamiliar with the reports, need to understand the logic behind the data. With the combination of the three layers, the technical personnel are able to understand the business considerations behind certain data, while the business personnel can get an insight into the logic behind the report information.*

This indicates that BI users do not necessarily understand how the data in any BI report are being constructed from within the firm's information systems, yet, when making business decisions, it is risky to rely on data whose source has not been clearly identified. To summarise, it was clear from the validation feedback that a BIP-Map framework is a valuable framework that can assist BI users to understand the entire business workflow and information manufacturing process (together with the metacontent) of any BI product before they make any important decisions for their organisation.

## CONCLUDING REMARKS AND CONTRIBUTIONS

This project applied a design science research paradigm (Hevner & Chatterjee 2010), and in this final section we show how the research fulfilled the seven guidelines for that paradigm as proposed by Hevner et al., (2004).

This study entailed the development of two artefacts. Referring to Hevner et al.'s Guideline 1 – design as an artefact, the initial artefact developed in this research was an integrated framework of Business Intelligence Product and Metacontent Map (BIP-Map) (i.e. a model). The second artefact was a BIP-Map informed prototype (i.e., an instantiation). The purposes of this research were to address the BI lineage problems and the individual-related data-quality issues and in this way we will be assisting BI users to better understand the BI products (Hevner et al.'s Guideline 2 – problem relevance). In achieving this, the designed artefacts have been strictly informed by, and incorporated with, three modelling and representation techniques; namely, the business process modelling notation (BPMN), the information product map (IP-Map), and the metacontent representation (Hevner et al.'s Guideline 5 – research rigour). The BIP-Map framework comprises three layers: the first layer adopts the business process model to describe both the business workflow and the participants involved. The second layer is an information product map which shows the information (i.e. BI product) manufacturing process and is constructed by expanding the first layer. The third layer represents the metacontent detail of the data validation stage at source system through to ETL, and the data warehousing stage.

In addition, the usefulness of the artefacts was validated via interviews by adhering to the interview methodology (Hevner et al.'s Guideline 3 – design evaluation and Guideline 5 – research rigour). The results of the framework validation will be used to refine the designed artefacts for future works. Such a build-and-evaluate loop will iterate several times before the final design artefacts are generated (Hevner et al.'s Guideline 6 – design as a search process). Moreover, this paper includes both the technical presentation and practical framing in terms of application in the online job advertising setting (Hevner et al.'s Guideline 7 – communication of research).

The designed artefacts are the major contribution of this research and as such they meet Guideline 4 – research contributions. The integration of the layers into one framework enables BI users to better understand BI products and so empowers them to respond promptly. In other words, this research explicitly addresses individual-related data-quality issues within a BI environment such as data believability, traceability, and accountability, issues that have not been addressed in earlier research but which can profoundly affect decision-making. It proposes an integrated framework to provide business workflow and lineage information which helps users understand the BI product and its manufacturing process. The prior IP-Map of Shankaranarayan et al. (2003) did not consider the business workflow and the metacontent from data validation through to the data warehousing stage that are related to the manufacturing of BI products in an organisation. This study, therefore, fills the research gap and it represents an extension of the foundational IP-Map research of Shankaranarayan et al. (2003). It also addresses the data-quality, data-stewardship, and business process management aspects of data governance by ensuring that data can be trusted and that people can be made accountable for any data quality issues. Furthermore, the BIP-Map could serve as a framework for BI practitioners to identify business and information phases that may create data quality problems. Technical personnel will be able to understand the business considerations that underlie certain data, while business personnel can obtain insights into the logic used in the creation of BI reports. In short, the BIP-Map framework builds confidence in the BI product by clearly outlining lineage information. Future research could apply the proposed BIP-Map framework to other business settings and test the utility of the framework.

## REFERENCES

Aguilar-Saven, R. S. (2004). Business Process Modelling: Review and Framework. *International Journal of Production Economics*, *90*(2), 129–149. doi:10.1016/S0925-5273(03)00102-6

Aiken, P., Gillenson, M. L., Zhang, X., & Rafner, D. (2011). Data Management and Data Administration: Assessing 25 Years of Practice. *Journal of Database Management*, *22*(3), 24–44. doi:10.4018/jdm.2011070102

Ballard, C., Farrell, D. M., Gupta, A., Mazuela, C., & Vohnik, S. (2006). *Dimensional Modelling: In a Business Intelligence Environment*. New York, United States: IBM.

Ballou, D. P., Wang, R. Y., Pazer, H., & Tayi, G. K. (1998). Modeling Information Manufacturing Systems to Determine Information Product Quality. *Management Science*, *44*(4), 462–484. doi:10.1287/mnsc.44.4.462

Bogza, R. M., & Zaharie, D. (2008). Business Intelligence as a Competitive Differentiator. In: *IEEE International Conference on Automation, Quality and Testing, Robotics*, Cluj-Napoca, Romania.

Carson, D., Gilmore, A., Gronhaug, K., & Perry, C. (2001). *Qualitative Research in Marketing*. London: Sage.

Delen, D., & Pratt, D. B. (2006). An Integrated and Intelligent DSS for Manufacturing Systems. *Expert Systems with Applications*, *30*(2), 325–336. doi:10.1016/j.eswa.2005.07.017

Dempsey, L., & Heery, R. (1998). Metadata: A Current View of Practice and Issues. *The Journal of Documentation*, *54*(2), 145–172. doi:10.1108/EUM0000000007164

Elbashir, M. Z., Collier, P., & Davern, M. (2008). Measuring the Effects of Business Intelligence Systems: The Relationship between Business Process and Organisational Performance. *International Journal of Accounting Information Systems*, *9*(3), 135–153. doi:10.1016/j.accinf.2008.03.001

Fisher, C., Lauria, E., Chengalur-Smith, S., & Wang, R. (2006). *Introduction to Information Quality*. USA: MIT Information Quality Program.

Foshay, N. (2005). *The Influence of End-User Metadata on User Attitudes toward, and Use of a Data Warehouse*. IBM, Somers.

Foshay, N., Mukherjee, A., & Taylor, A. (2007). Does Data Warehouse End-user Metadata Add Value? *Communications of the ACM*, *50*(11), 70–77. doi:10.1145/1297797.1297800

Gartner Press Release. (2013a). Gartner Executive Program Survey of More Than 2,000 CIOs Shows Digital Technologies Are Top Priorities in 2013, Stamford, Conn., Jan 16, 2013, Press Release. http://www.gartner.com/newsroom/id/2304615

Gartner Press Release. (2013b). Gartner Says Worldwide Business Intelligence Software Revenue to Grow 7 Percent in 2013, Stamford, Conn., Feb 19, 2013, Press Release. http://www.gartner.com/newsroom/id/2340216

Hevner, A., & Chatterjee, S. (2010). *Design Research in Information Systems: Theory and Practice*. New York: Springer. doi:10.1007/978-1-4419-5653-8

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *Management Information Systems Quarterly*, *28*(1), 75–105.

Howson, C. (2011). Despite Successes, BI Adoption is Flat, http://www.informationweek.com/news/software/bi/231902636. Accessed 17 June 2011.

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, *53*(1), 148–152. doi:10.1145/1629175.1629210

Negash, S. (2004). Business Intelligence. *Communications of the Association for Information Systems*, *13*, 177–195.

Ou, L., & Peng, H. (2006). Knowledge and Process Based Decision Support in Business Intelligence System. In: *Proceedings of the First International Multi-Symposiums on Computer and Computational Sciences*, Hanzhou, Zhejiang. doi:10.1109/IMSCCS.2006.236

Ponniah, P. (2001). *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals*. NY: Wiley. doi:10.1002/0471221627

Sen, A. (2002). Metadata Management: Past, Present and Future. *Decision Support Systems*, *37*(1), 151–173. doi:10.1016/S0167-9236(02)00208-7

Shankaranarayanan, G. (2005). Towards Implementing Total Data Quality Management in a Data Warehouse. *Journal of Information Technology Management*, *16*(1), 21–30.

Shankaranarayanan, G., & Cai, Y. (2006). Supporting Data Quality Management in Decision-Making. *Decision Support Systems*, *42*(1), 302–317. doi:10.1016/j.dss.2004.12.006

Shankaranarayanan, G., & Even, A. (2006). The Metadata Enigma. *Communications of the ACM*, *49*(2), 88–94. doi:10.1145/1113034.1113035

Shankaranarayanan, G., Wang, R. Y., & Ziad, M. (2000). IP-MAP: Representing the Manufacture of an Information Product. In: Proceedings of the 2000 International Conference on Information Quality, MIT, Boston.

Shankaranarayanan, G., Ziad, M., & Wang, R. Y. (2003). Managing Data Quality in Dynamic Decision Environments: An Information Product Approach. *Journal of Database Management*, *14*(4), 14–32. doi:10.4018/jdm.2003100102

Shi, Y., & Lu, X. (2010). The Role of Business Intelligence in Business Performance Management. In: *3rd International Conference on Information Management, Innovation Management and Industrial Engineering*, Kunming, China. doi:10.1109/ICIII.2010.522

Stock, D., & Winter, R. (2011). The Value of Business Metadata – Structuring the Benefits in a Business Intelligence Context. In A. d'Atri, M. Ferrara, J. F. George, & P. Spagnoletti (Eds.), *Information Technology and Innovation Trends in Organisations* (pp. 133–141). Heidelberg: Physica. doi:10.1007/978-3-7908-2632-6_16

Tallon, P. (2013). Corporate Governance of Big Data: Perspectives on Value, Risk, and Cost. *IEEE Computer Journal*, *46*(6), 32–38. doi:10.1109/MC.2013.155

Tan, W., Shen, W., Xu, L., Zhou, B., & Li, L. (2008). A Business Process Intelligence System for Enterprise Process Performance Management. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, *38*(6).

Vaduva, A., & Dittrich, K. R. (2002). Metadata Management for Data Warehousing: Between Vision and Reality. In: *International Symposium on Database Engineering & Applications*, Grenoble, France.

Wang, R. (1998). A Product Perspective on Total Data Quality Management. *Communications of the ACM*, *41*(2), 58–63. doi:10.1145/269012.269022

Wang, R., Lee, Y., Pipino, L., & Strong, D. (1998). Manage Your Information as a Product. *Sloan Management Review*, *39*(4), 95–105.

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, *12*, 5–34.

Wang, W., Ding, H., Dong, J., & Ren, C. (2006). A Comparison of Business Process Modelling Methods. In: *IEEE International Conference on Service Operations and Logistics, and Informatics Proceedings*, Shanghai, China. doi:10.1109/SOLI.2006.328910

Weske, M. (2007). *Business Process Management: Concepts, Languages, Architectures*. Berlin: Springer.

White, S. A. (2004). *Introduction to BPMN*. IBM Corporation.

Wixom, B., & Watson, H. (2010). The BI-Based Organisation. *International Journal of Business Intelligence Research*, *1*(1), 13–28. doi:10.4018/jbir.2010071702

Zhu, H., & Wang, R. Y. (2010). Information Quality Framework for Verifiable Intelligence Products, In Data Engineering. *Mining, Information and Intelligence, Springer*, *132*, 315–333.

## ENDNOTES

1. Based on information product concept, a BI product refers to any reporting information that is provided to the BI users for decision-making where information is treated much like the products found in the manufacturing environment of any industry (Wang 1998; Wang et al., 1998; Shankaranarayanan et al., 2000; Shankaranarayanan et al., 2003; Shankaranarayanan 2005).
2. In this paper, the terms metadata and meta-content are used interchangeably.
3. According to Foshay et al. (2007), lineage metadata "tells the user about the original source of data in the warehouse and describes what has been done to the data prior to being loaded into the warehouse. Lineage information answers the user's questions: Where did this data originate, and what has been done to it?"(p.72)
4. A detailed explanation of the BPMN building blocks is included in Appendix I.
5. A detailed explanation of the IP-Map building blocks is included in Appendix II.

*Chin-Hoong Chee is currently a Lecturer in Universiti Tunku Abdul Rahman (UTAR), Malaysia. He received his Master of Computer Science (by research) from the same university. His main research areas are in business intelligence and information quality. Before joining the academic, he worked as a programmer in a multinational software company.*

*William Yeoh is a tenured faculty member in School of Information and Business Analytics at Deakin University. He received his PhD from the University of South Australia and he has taught within Australia, Malaysia and Hong Kong. His research interests include business intelligence/ analytics, information quality and enterprise systems. He has published numerous research papers in peer-reviewed journals and conference proceedings. He is a frequent invited speaker in practitioners' conferences too.*

*Shijia Gao is a lecturer at the Faculty of Information Technology at Monash University. She received her PhD degree from the University of Queensland in 2010. Her research interests include business intelligence, decision support systems, decision theory, risk management, financial information systems, business process management, and knowledge management. Her research has appeared in Decision Support Systems, Journal of Decision Systems, Expert Systems with Applications, Journal of Knowledge Management, among others.*

*Gregory Richards is currently professor of performance management at the Telfer School of Management, University of Ottawa. Canada. His research interests include business intelligence, organisational learning, knowledge management and performance management. He is the Director of the IBM Centre for Business Analytics and Performance at the Telfer School of Management, and also manages a research cluster focused on public sector performance management. He has experience in industry having worked for Cognos Incorporated as well as in the public sector prior to his academic appointment.*

## APPENDIX 1

The elements of business process modelling notation (BPMN) can be categorised into four basic groups, namely, *Flow Objects*, *Connecting Objects*, *Swimlanes*, and *Artefacts* (White 2004). The function of each element in the four groups is briefly described below in Tables 1 to 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. The 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 categorise the activities to be conducted by different users into separate sections so that the interaction between all participants involved in a business process can be easily identified. The Artefacts group contains elements that are used to include additional information about how the activities are executed in a business process.

## APPENDIX 2

The Information Product Map (IP-MAP) is a modelling technique for representing the manufacturing processes of the information products available in an organisation or industry (Wang 1998; Wang et al., 1998; Shankaranarayanan et al., 2000; Shankaranarayanan 2005). With an IP-MAP, users are able to visualise 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 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 have gone through, and the flow from one place to another, based on different scenarios. The *Data Storage*, *Business Boundary* and *IS Boundary* blocks specify where the data are being stored or transferred throughout the organisation.

*Table 1. BPMN flow objects*

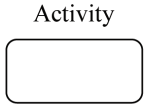
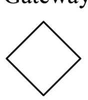
| Flow Objects | Function |
|---|---|
| Event  Start Intermediate End | - Event is something that happens during the execution of a business process and it is categorised as Start, Intermediate, and End. |
| Activity  | - Activity is the work performed in a business process. |
| Gateway  | - Gateway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

*Table 2. BPMN connecting objects*

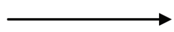
| Connecting Objects | Function |
|---|---|
| Sequence Flow  | - Sequence Flow shows the order of activities being performed in a business process. |
| Message Flow  | - Message Flow shows the communication of messages between two different parties. |
| Association  | - Association is used to indicate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

*Table 3. BPMN swimlanes*

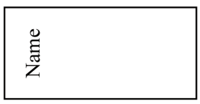
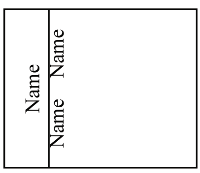
| Swimlanes | Function |
|---|---|
| Pool   | - A participant in a process is represented by a pool that acts as a graphical container for separating a set of activities. |
| Lane   | - The sub-section of a pool is called a lane and it separates the pool, either vertically or horizontally. |

*Table 4. BPMN artefacts*

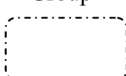
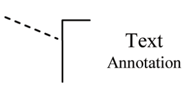
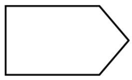
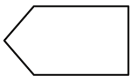
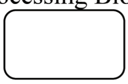
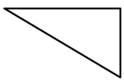
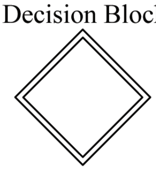
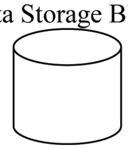
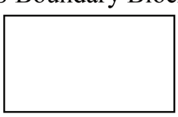
| Artefacts | Function |
|---|---|
| Data Object   | - Data Object shows the data that are captured or generated by an activity. |
| Group  | - Group categorises a set of activities together for the purpose of documentation or analysis. |
| Annotation   | - Annotation provides extra details about an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

*Table 5. IP-MAP building blocks*

| 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 directs users to the appropriate procedures. |
| Data Storage Block  | - Represents the capture of data in storage files or databases. |
| Business Boundary Block  | - Represents instances where the raw input data or component data are "handed over" by one business (organisational) 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. |