Journal of International Technology and Information Management

Volume 29 | Issue 2 | Article 4

2020

An Integrated View of Data: Application of Knowledge Modeling to Data Management

Sung-kwan Kim
University of Arkansas at Little Rock, sxkim@ualr.edu

Wenjun Wang
University of Arkansas at Little Rock, wwang@ualr.edu

Recommended Citation
Kim, Sung-kwan and Wang, Wenjun (2020) "An Integrated View of Data: Application of Knowledge Modeling to Data Management," *Journal of International Technology and Information Management*: Vol. 29: Iss. 2, Article 4.
DOI: https://doi.org/10.58729/1941-6679.1448
Available at: https://scholarworks.lib.csusb.edu/jitim/vol29/iss2/4

An Integrated View of Data: Application of Knowledge Modeling to Data Management

Sung-kwan Kim
(*University of Arkansas at Little Rock*)
Wenjun Wang
(*University of Arkansas at Little Rock*)

ABSTRACT

Data management has become an important challenge. Good data management requires an effective approach to collecting, storing, and accessing data across the enterprise. In this paper, a knowledge modeling approach to data management is introduced with an emphasis on data requirements analysis. A knowledge model can provide a high-level view of organizational data by specifying the structure and relationships of the knowledge contents used in business processes. The proposed knowledge modeling approach is business process oriented and decision oriented. The description of the knowledge contents in the model is based on ontological specification. The model is comprised of five elements: work product, work unit, producer, stage, and modeling language. The elements of the model and the modeling process are elaborated. The proposed modeling approach is applied to the vessel chartering process in a shipping company to demonstrate its application in real-world practices.

Keywords: Data Management, Data Requirement, Business Process, Knowledge Model, Ontology, Vessel Chartering

INTRODUCTION

In today's digital world, data has become a key word. With fierce global competition, business organizations are looking for new ways to gain competitive advantage. They need quality data that is timely and relevant for decision making and business operations.

Data provides business organizations with the capability to better understand the market and customer needs and translate them into products and services. Data guides business actions to cut costs, reduce time to market, and increase revenue opportunities. In fact, the ability to make quality business decisions is closely related to a company's capability to collect, analyze and use data properly. Therefore, many business organizations dedicate a tremendous amount of time and money in capturing, storing, processing and using data. A major challenge facing business organizations is how to manage the data in an effective and efficient manner.

As the importance of data increases, the need for an enterprise-wide approach to data management has also increased. Data management is the task of managing organizational data assets to meet the needs of organizations. Effective and efficient data management is crucial. The goals of data management are to create a long-term plan and strategy for data resources, analyze data requirements, identify opportunities for data use and sharing, develop data standards and policies, and control data quality and security (Fleckenstein & Fellows, 2018). Since the advent of the Internet and WWW, the amount of data has grown exponentially, and a new breed of data has emerged. These have brought many unfamiliar problems to the people in charge of corporate data assets. Today managers are facing new challenges in data management (Kim, 2011).

With the increased attention to data, there has been active research on the function of data management. A group of research emphasizes the data management framework. Examples include Data Management Association (DAMA)'s Data Management Body of Knowledge (DMBOK) model and Capability Maturity Model Integration (CMMI) Institute's Data Management Maturity (DMM) model. Another group of research emphasizes more specific areas of data management such as data quality, data security, data governance, data analytics, and data modeling (Fleckenstein & Fellows, 2018). Whichever approach they take, the first step in effective data management is to ask a question about "what to manage." Organizations assume that they fully understand what to manage and which data to capture. However, that may not be the case in many organizations. In data management, asking what to manage is a fundamental question. "Just because a set of data can be captured does not mean that it should be captured. Only data that can provide potentials to impact organizational behavior through their analysis should be included in an organizational data management frame" (Coronel & Morris, 2017, p. 654). Answering the question is closely tied to managing enterprise data requirement. The task of managing data requirements includes two activities: domain interpretation and model representation (Hadar & Soffer, 2006).

.

Domain interpretation is concerned with perceiving and interpreting a target domain and gathering requirements from the domain.

The second task, model representation, is concerned with translating the gathered requirements into a model. Even though the second task of model representation is a well-established discipline (e.g., conceptual modeling using entity-relationship model, unified modeling language, or object role model), the overall process of managing data requirements (particularly domain interpretation) lacks a methodical and systematic approach, which results in fragmented, disparate view of data due to model misrepresentation and model variations (Davis et al., 2004; Hadar & Soffer, 2006; Sommerville, 2016).

This paper introduces a knowledge-driven approach to data management, focusing on the analysis of data requirements. The approach uses a knowledge model to deal with data assets in various strategic, tactical and operational contexts, providing a more integrated view of data. The knowledge model allows users to determine and prioritize data requirements more methodically and systematically. The proposed knowledge modeling approach is business process oriented and decision oriented. It provides a single, enterprise-level description of data assets. Since it models on business processes, the knowledge model approach provides a more stable and enduring view of data assets. It creates an integrated structure of data assets and their management across the organization.

The rest of this paper is organized as follows. The next section reviews related studies on data management and knowledge modeling. In the following section, our knowledge modeling approach is proposed as a tool for data management. The model elements and the modeling process are elaborated. Then, the proposed knowledge modeling approach is applied to a vessel chartering process in a shipping company for demonstration. Finally, a conclusion is presented.

RELATED WORK

Data Management

In general, data is defined as raw facts about entities and events in the real world. Information is defined as data that has been processed. Processing represents the manipulation of data, such as averaging, totaling, grouping, sorting, and comparing. Therefore, information is more meaningful and valuable than raw data.

In this paper, we use the term “data” in the broad context to include both raw data and their processed form, i.e., information. In data-driven business environments,

.

data is critical inputs to business operations and decision making in various contexts. Data should be managed as a critical resource.

In general, management is a set of activities directed at an organization's resources with the aim of achieving organizational goals in an effective and efficient manner (Griffin, 2017). It includes the activities of planning, organizing, directing, and controlling organizational resources to achieve the desired goals. Accordingly, data management can be defined as the activity of planning, organizing and coordinating data to accomplish corporate objectives.

Traditionally, data management involves the policies, practices and technologies for acquiring, controlling, protecting, delivering and increasing the value of data (Pentek et al., 2017). More specifically, data management includes the activities for defining the data strategy; data management processes, standards, and measures; roles and responsibilities, data life cycle and architecture; and data model (Fleckenstein & Fellows, 2018). According to the Data Management Association (DAMA) Framework, data management is the collection of activities of developing and executing architectures, policies, practices and procedures that properly manage the full data lifecycle (DAMA, 2014). It involves many tasks such as developing a long-term plan, enforcing data standards and policies, determining data requirements, identifying opportunities for data use and sharing, and controlling data quality and security. Several data management frameworks have been developed. Two popular frameworks are CMMI's DMM model and DAMA's DMBOK model.

DMM focuses on assessing organization's maturity in given areas. This model divides data management into five high-level categories and one supporting process. The five high-level categories include data management strategy, data governance, data quality, data operations, and platform/architecture. The supporting process is a collection of activities that support adoption, execution, and improvement of data management practices, including measurement and analysis, process management, process quality assurance, risk management, and configuration management (Fleckenstein & Fellows, 2018).

DMBOK highlights data management domains. According to DMBOK2 Framework (DAMA, 2014), there are eleven data management knowledge areas: data governance, data architecture, data modeling and design, data storage and operation, data security, data integration and interoperability, document and content, reference and master data, data warehousing and business intelligence, metadata, and data quality.

.

Pentek et al. (2017) identifies eleven design areas in their reference model for data management, including business capabilities, data management capabilities, data strategy, performance management, people/roles/responsibilities, process and methods, data architecture, data lifecycle, data applications, data excellence, and business values. Other popular data management frameworks are MITRE's Data Management Domain Framework (MITRE-DMDF), Enterprise Data Management Council's Data Management Assessment Model (EDMC-DMAM), and Federal Enterprise Architecture Framework's Data Reference Model (FEAF-DRM) (Fleckenstein & Fellows, 2018).

Despite the increased managerial attention, a majority of companies underutilize or misuse the data they store. According to Dallemule & Davenport (2017), less than half of an organization's structured data is actively used in decision making, and less than 1% of its unstructured data is analyzed or used at all. They report many other data management problems including improper data access, data breaches, isolated data sets in silos, etc. There are many factors that inhibit successful data management such as lack of clear understanding of data requirements, misunderstanding of the data management concept, lack of strategic perspective, insufficient staff, technology-oriented mindset, and insufficient top management support (Kim, 2011). Poor data management results in fragmented and disconnected data. Consequently, data qualities are low. They are inconsistent, conflicting and confusing. These problems are critical concerns and must be resolved for effective and efficient data management function. In this paper, we particularly address the issue of lack of clear understanding of data requirements.

Collecting data and managing them are two different tasks. To be successful, organizations first need to think about how they will get solutions and benefits from the data they are collecting. Organizations assume that they understand what they need to know, though it may not be the case in reality. Often companies capture, store and organize data without clearly defining the business questions to which they wish to answer with the data. The right approach is to define the business questions, analyze the data requirements and collect the data to answer the business questions. Therefore, effective data management should begin with the understanding of an organization's need for data to support their business activities. To make truly data-driven business decisions, the entire process of data management must be driven by defining right business requirements and capturing the right data to meet the requirements.

There is a clear need for data planning that puts the organization's data needs first. Data planning creates a model of the business organization with its processes along with the data required. Formal data planning improves communication with users,

.

increases top management support, improves resource utilization, and develops an enterprise data architecture. The heart of data planning is the analysis of business processes and their data requirements (Goodhue et al., 1992). Identifying organizational data requirements is an essential part of data planning. Defining precisely what to build is one of the most challenging tasks in the development activities. Often the process is ambiguous and uncertain.

Many requirements engineering process models have been developed for requirements management. In their comparative study, Batra and Bhatnagar (2017) examine eight existing requirement engineering process models using the parameters such as requirements prioritization, feedback, support for reverse engineering, and risk management. Those models focus on requirement elicitation, documentation and validation rather than domain interpretation or organizational requirements understanding. They are oriented toward developing a product-specific project or application instead of identifying and synthesizing enterprise-wide data requirements.

There are also reference models concerned with the issue of requirements management on data level, such as CRISP-DM (Cross-Industry Standard Process for Data Mining), RAM (Requirements Abstraction Model), and REM (Requirements Engineering Reference Model). CRISP-DM (Chapman et al., 2000) is one of the most widely-used analytics models in data mining. It emphasizes in its business-understanding phase the importance of understanding the project objectives and requirements from a business perspective, and mentions in its data-understanding phase the need to evaluate whether the data acquired satisfies the requirements. However, it does not specify how to perform these tasks methodically. RAM (Gorschek & Wohlin, 2006) supports requirements management through the entire development process. It aims to refine the initially abstract and solution-independent requirements to software. REM (Geisberger et al., 2006) constitutes a methodic foundation for interdisciplinary development of requirements and system specification for embedded systems. One significant issue of these two reference models is that neither the requirements on different levels of abstraction nor the concretization of requirements is clear (Berkovich et al., 2012). To identify the potential business value and facilitate the collaboration between software engineers and data scientists, Altarturi et al. (2017) present a new requirement engineering model for big data systems, which sheds light on the importance and challenges to integrating organizational data and resources that may present in silos across the enterprise.

Despite considerable advancement in the field of requirements management, many issues are not yet satisfactorily solved. Requirements are still incorrectly identified,

.

frequently misunderstood, and vaguely expressed, leading to reworking and issues at the later stages in the life cycle (Sommerville, 2016). Requirements errors still produce most of the errors in the development projects, and most companies consider requirements analysis as very significant (Batra & Bhatnagar, 2017). A clear definition of business requirements with an integrated view of the organizational data is critical. However, it is not apparent how the task is best accomplished. There is a need for a systematic and business goal-oriented approach to data requirements analysis.

Knowledge Model

A knowledge model can be used as a tool for identifying, defining, and managing data requirements. A knowledge model is the representation of organizational knowledge resources. It describes the structure of knowledge entities, their relationships, usages and constraints in an enterprise. Knowledge modeling is a systematic way of analyzing knowledge and thereby identifying data requirements. Knowledge modeling approach is different from existing data analysis frameworks in that its analysis unit is knowledge. The focus of the analysis is on knowledge instead of data. Knowledge includes more than data and is a more complex concept. Many researchers and practitioners have defined it in their respective terms. Still there is no standard definition. Davenport & Prusak (1998, p. 5) define knowledge as “a mix of framed experiences values, contextual information and expert insight that provides a framework for evaluating and incorporating new experiences and information.” Wang & Noe (2010, p. 117) define knowledge as “information processed by individuals including ideas, facts, expertise, and judgement relevant for individual, team, and organizational performance”. Sveiby (1997) includes a capacity to act as an element. Using these explanations, we define knowledge as information combined with understanding, know-how, expertise and judgement learned through experience or study and actionable in a specific context. Business organizations can get a comprehensive view of data requirements by examining the knowledge they use. By analyzing knowledge in use, one can find both explicit and implicit data requirements. When a developer focuses his/her analysis on knowledge instead of data, even hidden data needs can be revealed.

Knowledge-driven analysis is more business process oriented than existing analysis approaches. As Ravesteyn and Batenburg (2010) pointed out, business process management influences many business and IT domains. This business-process oriented approach provides an effective way to our knowledge modeling process. We take the definition of business process used by Zoet et al. (2011), i.e., a set of linked procedures or activities which realizes a business goal within the context of organizational structure. Knowledge-driven approach is business process oriented

.

in that we create a knowledge model for a business process and focus on how the knowledge flows among activities to achieve business goals in the process.
Since organizational processes do not change frequently, process-based knowledge models are more stable.

Knowledge-driven analysis is also decision oriented in that modelers focus on the knowledge inputs for decision-making points in a process. Decision making is one of the fundamental processes for any business. A company has to make many decisions quickly and continuously. Business organizations are filled with decision making at various levels. Nearly all managerial activities revolve around decision making. For managers to make decisions, they need knowledge. An individual's problem-solving and decision-making capability is limited by the knowledge available. Having knowledge available to decision makers is crucial to improving individual and organizational performance. Therefore, the decision-oriented approach is a valid way of identifying knowledge requirements.

In knowledge modeling, it is popular to use ontology to specify knowledge contents. Ontological approach to modeling at the conceptual level gained popularity with its representation capability and expression power (Wand & Weber, 2004; Pinto et al., 2009). Ontology is the study of entities that exist in the world.
It is a formal, explicit specification of a shared conceptualization (Gruber, 1993). In its philosophical sense, it is the study of being. In the context of knowledge modeling, ontology means a specification of knowledge that can be designed for knowledge sharing and reuse (Pinto et al., 2009). The knowledge concept represents knowing about an entity which can be a person, thing, concept, event, or organization. Ontological description specifies conceptualizations of such entities formally (Gómez-Pérez, 2001). Ontological specification typically includes the description of properties, relationships, constraints, and behaviors of entities.
The properties describe the characteristics of an entity. The relationships explain how entities are related to each other. Constraints specify the rules governing the entities. Behaviors describe the actions the entities can take. Ontological study categorizes things that exist in the domain world. Ontology can be used as a means by which developers capture knowledge about a domain of interest. Ontology-based modeling supports a shared and common understanding of a domain and improves communication between the stakeholders by removing semantic heterogeneity.

Since ontology represents entities that exist conceptually or physically in reality, the ontological concepts remain constant as long as an enterprise stays in its business. Therefore, ontology-based knowledge modeling provides stability and reliability in representing and maintaining enterprise knowledge. The ontology-

.

based modeling is more enduring. Some of the popular ontology-based enterprise modeling methodologies are TOVE (Toronto Virtual Enterprise) Ontology, Enterprise Ontology, IDEF (Integrated Definition Methods) Ontologies, PIF (Process Interchange Format), NIST (National Institute of Standards and Technology), PSL (Process Specification Language) Ontology, CIMOSA (CIM Open System Architecture), PERA (Purdue Enterprise Reference Architecture), and GERAM (Generic Enterprise Reference Architecture and Methodology). KAON (Karlsruhe Ontology) and Semantic Web project is a framework for the development of ontology-based Semantic Web applications (Kayed et al., 2008). MethOntology, WebODE, and On-To-Knowledge are used for creating ontologies for information system development (Fonceca, 2007).

KNOWLEDGE MODELING METHOD

We propose a knowledge modeling method that can provide an integrated view of organizational knowledge contents. A method is defined as "*a way, technique or process for doing something*" (Bengsch et al., 2019, p. 243). Employing a good method is critical to building a reliable knowledge model. Unfortunately, there is no standard method available in knowledge modeling. In other conceptual modeling areas such as process modeling or data modeling, there are a few well-established and standardized methods, e.g., entity relationship model in data modeling, data flow diagram in process modeling, and unified modeling language in object-oriented modeling. Well-established standard methodologies often integrate best practices and provide an easy-to-use, yet expressive tool. Those methodologies provide a formal basis for designing and developing a knowledge model and facilitate the development process. In this section, we propose an ontology-based knowledge modeling approach to data management with an emphasis on data requirements analysis. Our approach renders an integrated view of organizational data.

Model Elements

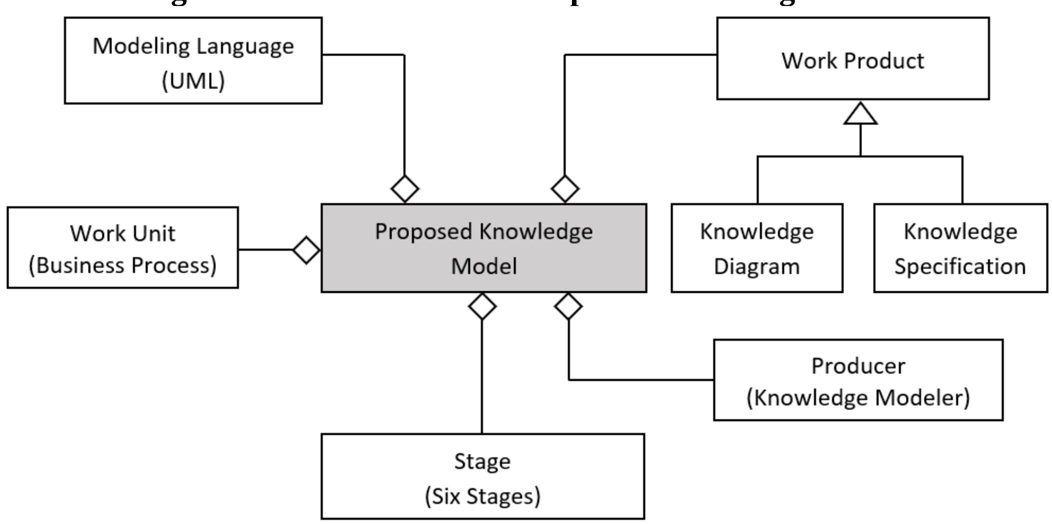
One of the popular metamodels at the conceptual level is the OPF (Open Process Framework) metamodel (Firesmith & Henderson-Sellers, 2000; Henderson-Seller, 2003). OPF defines five components a conceptual-level model should include: work product, producer, work unit, language, and stage. A work product is any valuable result of modeling process. A producer is the one who creates, evaluates, and maintains work products. A work unit is a functionally cohesive operation performed by a producer. A language is used to visualize and document work

.

products. A stage is an identifiable and manageable duration within the modeling process.

We adapt the OPF to our knowledge modeling approach. In our model, there are two work products: *knowledge diagram* and *knowledge specification*.
The knowledge diagram is the graphical representation of a knowledge model.
It provides a high-level view of organizational knowledge contents and their relationships. Knowledge specification is a textual model.
The text model is ontology-based and describes the details of knowledge contents including their structures, properties, behaviors, constraints, and managerial issues. The work unit in our model is a *business process*. A knowledge model is created for each business process. A producer in our model is any *knowledge modeler* who is responsible for developing the knowledge model. As for a language, our choice is UML (Unified Modeling Language). Since its standardization by OMG (Object Management Group) in 1997, UML has become an industry standard mechanism for visualizing, specifying, constructing, and documenting software systems. UML has proven to be effective for conceptual modeling because it has a very rich set of tools. The last component of the metamodel is a stage. Our knowledge model consists of *six stages*: business process selection, decision node identification, knowledge input analysis, knowledge diagram creation, knowledge specification, and model evaluation. Figure 1 shows the metamodel of the proposed knowledge model. UML notations are used to describe the metamodel. Each white rectangle represents a component of the model. A triangle notation represents a generalization and specialization (i.e., super/subtype) relationship. A diamond notation is for the aggregation (i.e., component/assembly) relationship.

Figure 1. Metamodel of the Proposed Knowledge Model

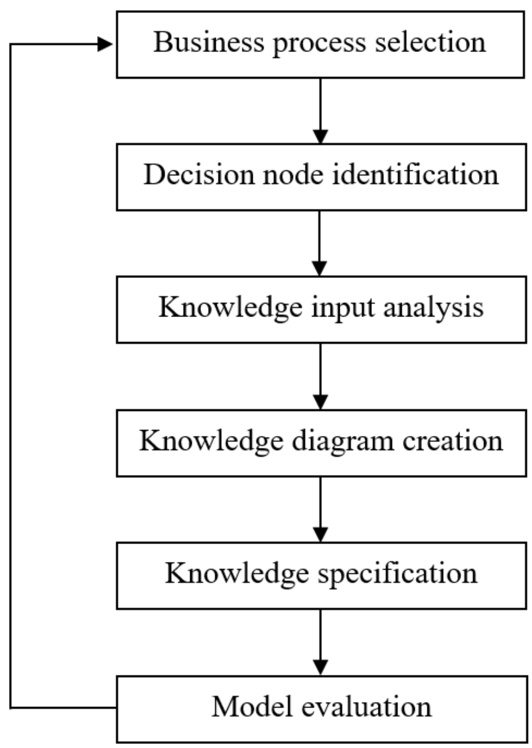


.

Knowledge Modeling Process

As described in the previous section, our knowledge model consists of a knowledge modeler, a specific business process, six stages, the UML, and two work products. Knowledge modeling is basically a collection of activities in which the knowledge modeler selects a business process, and go through the six stages to create the work products. In this subsection, we elaborate each stage as well as the two work products. Figure 2 illustrates the knowledge modeling process.

Figure 2. Knowledge Modeling Stages



Business Process Selection

The first stage in our knowledge modeling is to select a business process.
A business process is a set of linked procedures or activities which realize a business goal within the context of organizational structure (Zoet et al., 2011).

.

A business process is used to coordinate and organize work activities, data, information, and knowledge to produce a valuable product or service. Identifying a business process defines a conceptual framework for which a knowledge model is being created. Each process requires various knowledge to accomplish its mission. A knowledge model is created for one process. Therefore, multiple knowledge models are needed since a business organization uses more than one process.
In general, when multiple models are created, they are integrated into a single, integrated model.

Decision Node Identification

The next stage is to identify decision nodes in the business process. As indicated earlier, our knowledge modeling is decision oriented. A decision node is where a decision is made in the chosen process. A business process usually contains multiple decision nodes. To make quality decisions, decision makers need knowledge that provides context-specific intelligence. Having the relevant knowledge (both in quality and quantity) available to decision makers is crucial to improving individual and organizational performance. Therefore, the decision-oriented approach is an effective way of identifying knowledge requirements. Decision-oriented models can be used to explain how and why the process proceeds. Thus, a decision-oriented modeling paradigm is considered to be appropriate for knowledge modeling process (Rolland et al., 1999).

Knowledge-Input Analysis for Decision Nodes

The next stage is to analyze the knowledge required for each decision node. Knowledge is a collection of related data, expertise, and skills necessary for decision making in a specific context. Knowledge for each decision node must be identified. Each decision node requires various types of knowledge inputs. This task of identifying knowledge inputs should be done with the domain experts. The emphasis should be put on the identification of the knowledge required for decision making, instead of the knowledge currently being used or available. The goal is to specify knowledge requirements no matter whether the knowledge is currently available or not. All the necessary knowledge inputs for each decision node in the process need to be analyzed and identified.

Knowledge Diagram Creation

The next stage is to create a knowledge diagram. It is used to visualize and organize all the knowledge contents identified. It is one of the two work products of our

.

knowledge model. A knowledge diagram is a graphical representation of knowledge contents. Specifically, the diagram shows a set of knowledge units with their ontological concepts and their relationships. Each knowledge unit represents a logically related group of knowledge needed for decision making at a specified decision node. The use of ontology promotes the knowledge reuse and sharing across multiple domains. To represent the ontological concepts, we apply the Ontology Generic model by Fox (1992). The relationships describe how knowledge units are related to each other. The typical relationships are generalization and specialization (e.g., student and undergraduate student), association (e.g., doctor and patient), and aggregation (e.g., car and engine). Any modeling languages that are able to express entities and their relationships can be used to create a knowledge diagram. We use UML (Unified Modeling Language). UML's Class Diagram provides an excellent way for creating a knowledge diagram. An example of knowledge diagram is presented in the demonstration section.

Knowledge Specification

The next stage in our knowledge modeling process is knowledge content specification. All the knowledge contents identified in the previous stages are specified in detail. The formality of the documentation determines how the specification removes the ambiguity of natural languages. There are levels of formalization from the highly informal to the rigorously formal (Kayed et al., 2008). Highly informal specification may use a natural language to express the knowledge contents. It is flexible and easy to understand for human beings. This approach has full representation power but is not efficient. It is not precise and therefore includes a lot of ambiguities. The rigorously formal approach specifies terms in formal semantics (e.g., logic-based language) so that their properties are well understood with the least ambiguity. However, this approach lacks the rich representation power of natural languages (Uschold & Gruninger, 2004; Kayed et al., 2008). As the formality increases, ambiguity is reduced. For each level of formalization, there is a tradeoff between the representational power of a language and the efficiency of the reasoning engines. We take an ontological specification. The knowledge concept represents knowing about an entity which can be a person, thing, concept, event, or organization. Ontological description specifies conceptualizations of such entities formally (Gómez-Pérez, 2001). Ontological specification typically includes the descriptions of properties, relationships, constraints, and behaviors of entities. Properties describe the characteristics of an entity. Relationships explain how entities are associated with each other. Constraints specify the rules governing the entities, and behaviors describe the actions the entities can take. This ontological approach provides a formality without losing expressive power. It maintains a certain level of formalization.

.

At the same time, it is appropriate for human communication purposes.

A knowledge specification for each knowledge unit in our knowledge modeling approach is composed of five components: structured components, unstructured components, relationships, constraints, and managerial components. Structured components represent the explicit part of a knowledge unit, including all data that can be clearly defined, expressed, and communicated. They are the intrinsic properties of a knowledge unit and situation independent. For example, the knowledge unit, VESSEL, has properties such as name, class, dimensions, speed, tonnage, cargo holding capacity, and fuel consumption. Any composite properties are broken down into subcomponents. For example, the dimension of VESSEL is a composite property. It can be broken down into length, breadth, and depth.

Unstructured components represent the implicit or behavioral part of a knowledge unit. They include actionable operations, expertise and experiential skills. They are implicit and context-specific characteristics of the knowledge unit. They can be learned from education/training, experiences, or trial and errors. Since they are implicit, the unstructured components of a knowledge unit are difficult to define and hard to communicate. Actually, these components are what distinguish knowledge from mere data or information.

Any constraints on the use of specific knowledge contents should be described. Constraints are important because they help ensure data quality. For example, to deal with the knowledge about cargo demand and supply, a certain level of experience and training is required. This constraint ensures only qualified people can manage the knowledge contents.

Managerial components describe how a knowledge unit is managed. From the managerial perspective, the answers to the following questions about a knowledge unit are critical for maintaining an accurate and reliable knowledge model: What is the managerial impact of the knowledge? How critical is the knowledge to the organization's missions and objectives - strategic, tactical or operational? Where is it originated? Who produces the knowledge? Who consumes it? Is it acquired from internal sources or external sources? In which format? Who maintains it? Is the knowledge current? How frequently is the knowledge evaluated and updated? All of these questions are essential for maintaining the relevance and accuracy of the knowledge model. An example of knowledge specification is presented in the demonstration section.

.

Model Evaluation

The last stage in our knowledge modeling process is to evaluate the model. A good model represents the domain with accuracy and completeness. It should be validated rigorously (Shanks et al., 2003). Developers want to see if their model appropriately represents the corresponding reality and creates value. Kim (2014) proposed a model of four dimensions for evaluating a knowledge model: validation, representation, applicability and management. Validation and representation are used to evaluate the effectiveness of a model. Applicability and management are used to evaluate the efficiency of a model.

The validation dimension evaluates the model's correspondence with the problem domain for which the model is being constructed. Three constructs support the validation dimension: validity, completeness and accuracy. Validity confirms that the knowledge model corresponds to the domain that it is supposed to represent. Cooper & Schindler (2006) defines validity as the extent to which a test measures what it actually wishes to measure. Validity is achieved through a final review of the model with domain users to ensure that the model is an accurate representation of organizational knowledge. Completeness means that the model contains all the constructs and definitions that are correct and relevant within the domain. A model is complete if it covers all elements in the target domain. Accuracy measures the precision of a model's conformity to the domain. It checks how precisely the model covers the elements in the domain. If the model represents the target domain well and accurately, it is functionally complete. (Wand & Webber, 2004; Siau & Rossi, 2011).

The representation dimension assesses the syntactical aspects of the model.
It measures syntactical correctness, consistency, conciseness and richness of the modeling language. Syntactical correctness is an important criterion for evaluating representation quality. If all notations and their usages in the model conform to the grammar and constraints of the language, the model is said to be syntactically correct. Consistency refers to whether it is possible to obtain the same outcomes when valid inputs are given. A model is consistent if it does not produce conflicting results when valid inputs are given. Conciseness evaluates if a model does not store any unnecessary or useless definitions. For a model to be concise, there should be no redundancies in notations, definitions and axioms (Fox & Gruninger, 1998). Richness measures the expressive power of the model. A model should provide sufficient semantic concepts so that it can describe all relevant aspects of the problem domain.

.

The applicability dimension evaluates how useful the model is. The assessment focuses on evaluating the generality, usefulness, and usability of the model from the user point of view (Gómez-Pérez, 2001). Generality evaluates how well a model is applicable across different domains. The more general the model is, the more compatible and shareable it is with other domains (Fox & Gruninger, 1998). Usefulness evaluates the relevancy of the model to the user context. The more relevant the model is, the more applicable it is. Usability is a concept that assesses how easily a user can interact with the model. It also means how easy it is for users to accomplish basic tasks when the model is used for the first time (Nielsen, 2012).

Most evaluation methods focus on the technical aspects of the model. There is a lack of emphasis on the managerial aspects of the model. Some of the managerial issues to be addressed are who will evaluate the model, when and how the evaluation process will be performed, and how the evaluation results will be accepted (Dieng et al., 1999). The concepts supporting this dimension include maintainability, reusability, mode and frequency. Maintainability determines the ease with which the maintenance can be carried out. As knowledge evolves and changes inevitably, a knowledge model must be appropriately maintained. Regular audit and evaluation are necessary to maintain an effective model. Reusability refers to taking components of one product in order to facilitate the development of a different product with different functionality. This concept measures the extent to which all or part of the model can be reused in different model development. The mode and frequency of evaluation should be also considered. The evaluation mode can be informal or formal. A formal evaluation is performed by using a standard methodology while an informal evaluation can be conducted by a development team and users. A focus group of experts or a walkthrough by users and developers are excellent ways of evaluating a model. Determining the frequency of evaluation and the establishment of a mechanism that links the feedback to the model are also critical issues in the model evaluation.

It is noted that our knowledge modeling process is not a one-time job. It is an iterative process. Not only does the modeler have to check back and forth between stages, but he/she may also have to get back to the business process selection stage from the model evaluation stage to ensure the model satisfies the original business process.

.

DEMONSTRATION

In this section, we apply the knowledge modeling approach to a vessel chartering process in a Korean marine shipping company for demonstration. The company is affiliated with one of the major cement manufacturing companies in Korea. The company specializes in transporting cement products to domestic ports. It owns a dozen of cement tankers of various sizes for bulk cement transportation. One of the authors of this paper worked for the company as a manager in charge of vessel operations including vessel chartering. He has extensive experiences and is very knowledgeable about the process. Most shipping companies own less tonnage than actually needed for their cargos and depend on chartering vessels whenever necessary.

A voyage charter is taken as an example. A voyage charter is a contract in which a vessel is hired for a voyage between a loading port and a discharging port. The charterer pays the vessel owner based on the cargo amounts or lump-sum basis. The owner is responsible for the port fees, fuel costs, and crew costs. Chartering is a complex process and requires a depth of knowledge. This example is not designed to accurately and completely represent the sophisticated process of chartering. Instead, it illustrates how the proposed approach can be applied to the real-world practice. For that purpose, the process has been simplified.

Process Selection and Decision-Node Identification

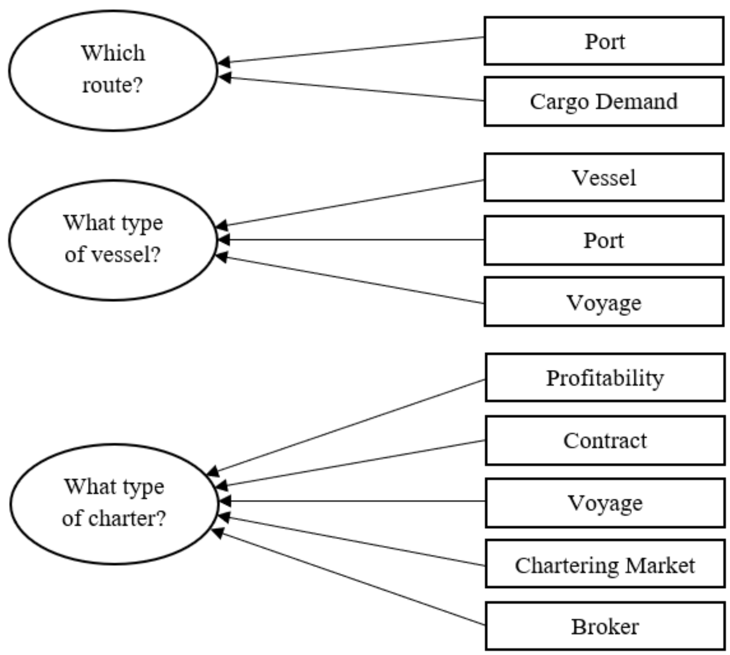
The first two stages in the knowledge modeling process are to select a business process and identify decision nodes in the process. The business process selected in this example is the chartering process that includes multiple decision nodes. Any decision nodes in the process need to be identified. They will be used to analyze knowledge inputs. Three decision nodes have been identified. The first one is to decide which route to choose. The charterer should choose where the cargo should be loaded and shipped (e.g., loading and discharging ports). The second decision node involves vessels. The charterer should decide on the type and size of a vessel to be hired. The third one involves a charter. The charter type should be decided.

Knowledge-Input Analysis for Decision Nodes

A decision maker needs many knowledge inputs to make decisions. By analyzing the knowledge inputs needed for decision making, developers can identify knowledge contents required. For example, to decide on the route, the charterer needs to know the cargo demand such as inventory in a cement silo, daily consumption rate, and construction business in the cities where bulk cement is supplied.

The charterer also needs to know about ports, such as loading/discharging facilities, water level, berth, and tidal change. To decide on vessel type to hire, the charterer needs to know about the voyage, the ports and the vessels. Particularly, the specifications of the knowledge about vessels, such as the type, tonnage, dimensions, engine type, speed, and fuel consumption, are important. To decide on the charter type, the charterer needs to know about the hire rate, terms and conditions, type of contract, voyage profitability, responsibilities, risks, liabilities, and so on. The charterer also needs knowledge about the chartering market including the brokers who arrange the chartering transaction. A knowledge input may be involved in one or more decision nodes. Figure 3 illustrates the three decision nodes and their knowledge inputs.

Figure 3. Knowledge Input Analysis for Decision Nodes

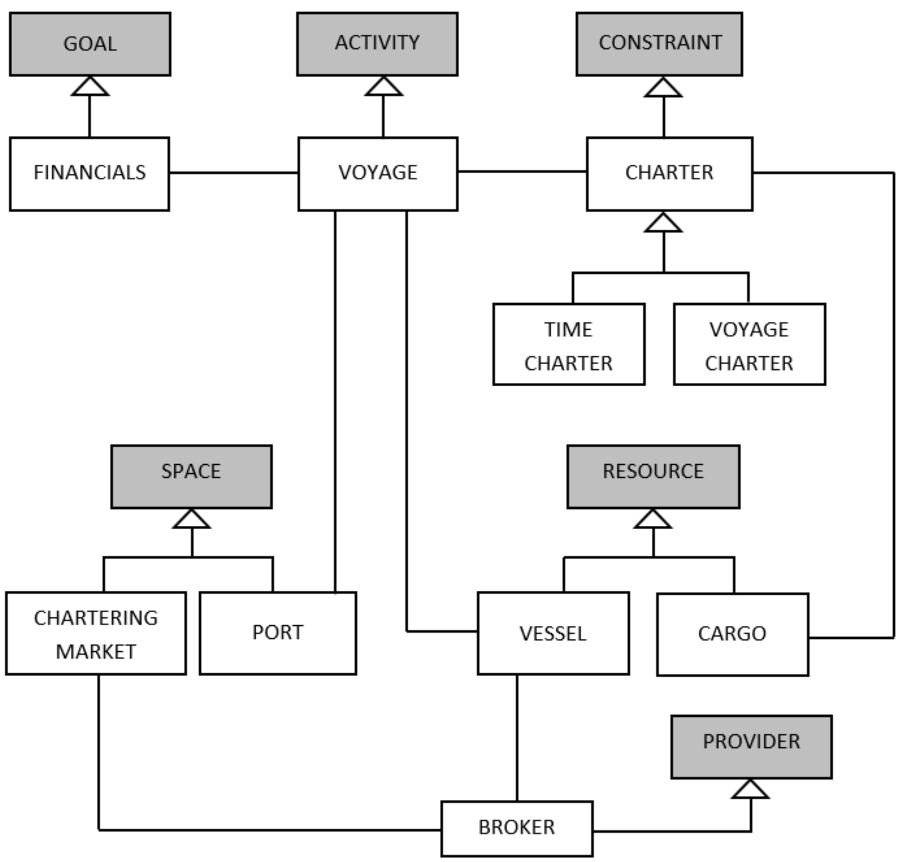


.

Knowledge Diagram Creation

The first work product of our knowledge modeling is a knowledge diagram. The knowledge diagram graphically shows all knowledge units identified in the previous stages. It also describes the relationships between them at a high level, providing a bird's eye view of the knowledge contents for the selected business process. The typical relationships between knowledge units are generalization and specialization, aggregation and association. A modified UML class diagram is used for creating a knowledge diagram, which provides an effective way of depicting knowledge contents. Figure 4 shows a portion of the knowledge diagram for the vessel chartering process. In the figure, a white rectangle represents a knowledge unit. A shaded rectangle represents an ontological concept. A triangle notation represents a generalization and specialization relationship, and a straight line represents an association relationship.

Figure 4. Knowledge Diagram



.

Knowledge Specification

The next stage in our knowledge modeling is to create knowledge specifications. The knowledge diagram shows the knowledge inputs and their relationships for the chosen process. However, it does not provide information about their internal structures and properties. A modeler needs to describe the structure and behaviors of each knowledge unit in the diagram. We take an ontological approach to the knowledge specification. Ontological specification typically includes the descriptions of properties, relationships, constraints, and behaviors of entities, which are the knowledge units in our knowledge model. The knowledge specification incorporates five components: structured components, unstructured components, relationships, constraints, and managerial components.

As explained in the previous section, structured components include the explicit part of each knowledge unit. The attributes of a knowledge unit belong to this group. If there are any composite attributes, they should be broken down into smaller components. A vessel's dimension is an example of a composite attribute. It can be further broken down into length, breath, and depth. Table 1 provides examples for the structured components of three knowledge units.

Table 1. Examples of Structured Components

Knowledge Units	Structured Components
Vessel	name, class, flag, type, year built, builder, size, dimension, cargo capacity, speed, engine type, fuel consumption, etc.
Voyage	voyage number, cargo type, cargo quantity, freight rate, loading port, discharging port, departure and arrival time, despatch, demurrage, etc.
Charter	type, hire rate, port of delivery, time of delivery, broker, charter party, trade limitation, duration, responsibilities, etc.

Unstructured components represent implicit and context-specific characteristics of a knowledge unit. Sea-worthiness evaluation is an example of the unstructured component of the knowledge unit, VESSEL. At the time the contract is made, the vessel must be fit to deal with ordinarily anticipated perils of the sea and incidental risks on a voyage. This task is complex and requires sophisticated knowledge such as assessment of the vessel's condition and suitability for the planned voyage. Risk assessment and liability assessment provide another two examples. Risk assessment

.

includes the analysis of safety of ports/berth and delays during the charter due to stevedore strikes or war. Liability assessment involves the analysis of obligations from the loss or damage to cargo and damage to hull by cargo. These types of knowledge are based on the evaluator's experience and expertise. By analyzing them, the evaluator may reveal data requirements that are not found in the explicitly structured data.

Any constraints on handling a knowledge unit are also specified. For example, the person who handles the knowledge unit, VESSEL, must have a specific number of years' experience and training as a ship officer or engineer. If other qualifications (e.g., certificate) are required, they must be specified as well.

Managerial components of the knowledge unit are also described, including who owns the knowledge. For example, the knowledge unit, VESSEL, is owned by the marine affairs department. The knowledge unit, VOYAGE, is owned by the operation department. Format of the knowledge unit is specified as well. It can be in the form of document (e.g., manuals, policies, or any other reports) and can reside in the internal or external databases. The acquisition mode is documented. It can be produced internally by the employees or externally by outside experts. Managerial criticality is also documented. For example, the knowledge about VOYAGE has tactical implications. The knowledge about VESSEL may be operational. Finally, there are maintenance issues. Knowledge evolves and must be continuously updated. Table 2 provides an example of partially filled knowledge documentation for the VOYAGE knowledge unit.

Table 2. Example of Knowledge Unit (KU) Specification

KU Name	**VOYAGE**
Ontology Concept	ACTIVITY
Relationships	CHARTER (Association), PROFITABILITY (Association)
Description	This unit includes the knowledge for a voyage.

.

Structured components (Properties)	- Voyage number - Vessel name - Cargo type/quantity - Loading/discharging ports - Freight rate - Place and time of delivery - Agency fee - Port charges - Fuel costs - Demurrage and despatch (Dem/Des)
Unstructured components (Behaviors)	Calculate charter base Calculate Dem/Des Assess risks - Port safety - Delay during the charter - Cancellation of charter Assess liabilities - Loss or damage to cargo - Damage to hull by cargo - Measure of damages
Constraints	- At least 5 years of experience in chartering - Completion of vessel chartering training program
Managerial components	- Owner: operation department - Acquisition: internal - Format: internal database - Criticality: tactical - Maintenance: when, how, by whom, how often, which mode?

CONCLUSION AND FUTURE WORK

In this paper, we propose an ontology-based knowledge modeling approach as a data management tool for analyzing data requirements. Our approach provides a high-level, integrated view of organizational data by specifying the structure and relationships of knowledge contents used in business processes. In our model, the knowledge modelers use UML as the modeling language and go through six stages to create a knowledge diagram and associated knowledge specifications for each business process.

.

Our approach has three major strengths. First, our model focuses on the data requirements at the knowledge level, even hidden data needs are disclosed. It models what is required instead of what is currently available. Therefore, future knowledge/data requirements can be planned and managed appropriately. If a business organization does not have the required knowledge, it is expected to trigger a system to acquire it. Data management should work in the knowledge management ecosystem because what a company really needs is knowledge, not mere data. Data themselves do not add values. Second, the proposed knowledge modeling highlights decision nodes in business processes. Focusing on critical decision nodes avoids distraction that can result from a too detailed or unnecessary analysis. In addition, since our knowledge model is ontology-based and decision/business-process oriented, our approach renders a more stable and enduring view of the organizational data assets. Finally, our approach examines the data requirements across the enterprise, and provides an integrated view of organizational data. It helps facilitate the integration of organizational data/resources and enterprise-wide collaboration. Moreover, our approach is domain independent. It can be applied to different domains in business.

A modeling method can be validated by reviews via focus group or problem solving in the focal domain (Shanks et al., 2003). As demonstrated in this paper, our knowledge modeling approach has been applied to the vessel chartering process in a real-life shipping company. It works well. The vessel operations department of the company never adopted any model-driven approach to data management before. Using our knowledge modeling approach, the department is now able to systematically understand, elicit and visualize the data requirements (including their structures and relationships). This study provides insights and implications for both data-management research and real-world practice.

There are some limitations of this research and interesting directions for the future work. Although the proposed knowledge modeling approach has been validated in the case study, more empirical and more rigorous evaluations are expected. The vessel chartering process is a small part of the entire vessel operations process. It is desirable to evaluate the model by applying it to the full context of the vessel operations. In the full context, multiple processes (e.g., vessel operating, voyage management, and crew management) interact and involve multiple departments in the shipping company. Since a knowledge model is created for each process, multiple knowledge models will be created. Integrating multiple models is a daunting task. There are a lot of merging and mapping issues to be resolved. Moreover, with multiple departments involved, maintaining the model with effective knowledge sharing mechanism is another important challenge. How to maintain the model integrity and reduce redundancy should be carefully planned.

.

This study is the first part of the larger comprehensive study. Improvement in the modeling method will be made with more real-world applications. A lot of interesting work can follow.

REFERENCES

Altarturi, H., Ng, K. Y., Ninggal, M., Nazri, A., & Ghani, A. (2017). A requirement engineering model for Big Data software. *In Proceedings of the 2017 IEEE conference on Big Data and Analytics (ICBDA).* Beijing, China.

Batra, M., & Bhatnagar, A. (2017). A comparative study of requirements engineering process model. *International Journal of Advanced Research in Computer Science*. 8(3), 740-745.

Bengsch, J., Steenbergen, M., & Ravesteyn, P. (2019). Fit for purpose enterprise architecture. *In Proceedings of the 32nd BLED Conference: Humanizing Technology for a Sustainable society*. Bled, Slovenia.

Berkovich, M, Leimeister, J. M., Hoffmann, A. & Krcmar, H. (2012). A requirements data model for product service systems. *Requirements Engineering*. 19(2), 161-186.

Calvanese, D., De Giacomo, G., & Lenzerini, M. (2001, July). Ontology of integration and integration of ontologies. In *Proceedings of the 14th International Workshop on Description Logics,* Stanford, CA.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*. The CRISP-DM consortium.

Cooper, D., & Schindler, P. S. (2006). *Business Research Methods.* New York, NY: McGraw-Hill/Irwin.

Coronel, C., & Morris, S. (2017). *Database system: Design, Implementation and Management*. Boston, MA: Cengage Learning.

Dallemule, L., & Davenport, T. (2017). What's your data strategy? *Harvard Business Review*, *May-June*, 113-121.

Data Management Association (DAMA) (2014). *The Data Management Body of*

.

Knowledge 2 (DMBOK2) Framework. Retrieved March 6, 2014 from https://www.dama.org/sites/ default/files/download/DAMA-DMBOK2-Framework-V2-20140317-FINAL.pdf

Davenport, T. H., & Prusak, L. (1998). *Working Knowledge: How Organizations Manage What They Know*. Boston, MA: Harvard Business School Press.

Davis, I, Green, P, Rosemann, M, & Gallo, S. (2004). Conceptual modeling-what and why in current practice. Atzeni, P. et al. (Eds.): ER 2004, LNCS 3288, 32-42. Springer-Verlag: Berlin Heidelberg.

Dieng, R., Corby, O., Giboin, A., & Ribière, M. (1999). Methods and tools for corporate knowledge management. *International Journal of Human-Computer Studies*. 51(3), 567-598.

Eriksson, H., & Penker, M (2000). *Business Modeling with UML: Business Patterns at Work*, New York: NY: John Wiley & Sons, Inc.

Firesmith, D., & Henderson-Sellers, B. (2002). *The OPEN Process Framework: An Introduction*. Addison-Weskey, Harlow, Herts, UK.

Fleckenstein, M., & Fellows, L. (2018). *Modern Data Strategy*. Cham, Switzerland: Springer International Publishing AG.

Fonceca, F. (2007). The double role of ontologies in information science research. *Journal of the American Society for Information Science and Technology*, 58(6), 786-793.

Fox, M. (1992). The TOVE project: Toward a common-sense model for the enterprise. *In Proceedings of International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems.* Paderborn, Germany.

Fox, M., & Gruninger, M. (1998). Enterprise modeling, *AI Magazine*, 19(3), 109-121.

Geisberger, E., Broy, M., Berenbach, B., Kazmeier, J., Paulish, D. & Rudorfer, A. (2006). *Requirements engineering reference model (REM)*. Technische Universität München, München.

Gembasu, Cristina. V. (2012). BPMN vs. activity diagram for business process modeling. *Accounting and Management Information Systems*, 11(4), 637-651.

.

Gómez-Pérez, A. (2001). Evaluation of ontologies. *International Journal of Intelligence Systems*,
16(3), 391-409.

Goodhue, D. L., Kirsch, L. J., Quillard, J. A., & Wybo, M. D. (1992). Strategic data planning: Lessons from the field. *MIS Quarterly*, 16(1), 11-34.

Gorschek, T., & Wohlin, C. (2006). Requirements abstraction model. *Requirements Engineering*. 11(1), 79-101.

Griffin, Ricky W. (2017). *Management*. Boston, MA: Cengage Learning.

Gruber, T. R. (1992). Toward principles for the design of ontologies used for knowledge sharing. *In Proceedings of the International Workshop on Formal Ontology*, Padova, Italy.

Henderson-Seller, B. (2003). Method engineering for OO systems development. *Communications of the ACM*, 46(10), 73-78.

Hadar, I & Soffer, P. (2006). Variations in conceptual modeling: classification and ontological analysis. *Journal of the Association for Information Systems*, 7(8), 568-592.

Kayed, A., Hirzallah, N., Al Shalabi, L. A., & Najjar, M. (2008). Building ontological relationships: A new approach. *Journal of the American Society for Information Science and Technology*, 59(11), 1801-1809.

Kim, S. (2011). A survey of Data Resource Management, In *Proceedings of International Association of Business and Public Administration*. Dallas, TX.

Kim, S. (2014). A managerial perspective on enterprise knowledge model evaluation. *International Journal of Business and Public Administration*, 11(2), 120-130.

Kim, S., Felan, J., & Kang, M. (2011). An ontological approach to enterprise knowledge modeling in a shipping company. *International Journal of Knowledge Management*, 7(4), 70-84.

Maedche, A., Motik, B., Stojanovic, L., Studer, R., & Volz, R. (2003). Ontologies

.

for enterprise knowledge management. *IEEE Intelligent systems*, 28(2), 26-33.

Nielsen, J. (2012). *Usability 101 introduction to usability*, Retrieved May 10, 2014 from http://www.nngroup.com/articles/usability-101-introduction-to-usability.

Pentek, T., Legner, C., & Otto, B. (2017). Towards a reference model for data management in the digital economy. In *Proceedings of the 12th International Conference on Design Science Research in Information Systems and Technology*. Karlsruhe, Germany.

Pinto F. M, Marquez, A., & Santos, M. F. (2009). Ontology-supported database marketing. *Database Marketing & Customer Strategy Management*, 16 (2), 76-91.

Pottinger, R. A., & Bernstein, P. A. (2003). Merging models based on given correspondences. *In Proceedings of the 29th Very Large Database (VLDB) Conference*. Berlin, Germany.

Ravesteyn, P. & Batenburg, R. (2010). Surveying the critical success factors of BPM-system implementation. *Business Process Management Journal,* 16(3), 492-507.

Rolland, C., Nurcan, S., & Grosz, G. (1999). Enterprise knowledge development: the process view. *Information and Management,* 36(3), 165-184.

Shanks, G., Tansley, E., and Weber, R. (2003). Using ontology to validate conceptual models. *Communications of the ACM*, 46(10), 85-89.

Siau, K., & Rossi, M. (2011). Evaluation techniques for system analysis and design modeling methods – a review and comparative analysis. *Information Systems Journal*, 21(3), 249-268.

Sommerville, I. (2016). *Software Engineering*. 10th Edition, England: Pearson Education Limited.

Sveiby, K.E. (1997). *The New Organizational Wealth: Managing and Measuring Knowledge Based Assets*, San Francisco, CA: Berett-Koehler Publishers.
Uschold, M., & Gruninger, M. (2004). Ontologies and semantic for semantic connectivity. *SIGMOD Record*, 33(4), 58-64.

Wang, S., & Noe, R. A. (2010). Knowledge sharing: A review and directions for future research. *Human Resource Management Review*, 20(2), 115-131.

.

Wand, Y., & Weber, R. (2004). Reflection: Ontology in information systems. *Journal of Database Management*, 15(2), 3-6.

Zoet, M., Versendaal, J., Ravesteyn, P., & Welke, R. (2011). Alignment of business process management and business rules. *In Proceedings of the 19th European Conference on Information Systems*. Helsinki, Finland.

.