Business Intelligence System Using Goal-Ontology Approach: A Case Study in Universiti Utara Malaysia

Azman Ta'a and Mohd Syazwan Abdullah

Universiti Utara Malaysia, Malaysia

azman@uum.edu.my, syazwan@uum.edu.my

Business Intelligence (BI) system design involves several tasks such as defining DW requirements, modeling DW structure and specifying ETL processes operations, and these have been studied and practiced for many years. However, the common design-related problems such as defining user requirements and deriving data integration and transformation activities are still far from being resolved due to the ambiguity of business requirements and the complexity of DW and ETL process, and that was the fundamental issues of conflicts in heterogeneous information sharing environments. Current approaches that are based on existing software requirement methods still have limitations on reconciliation the business semantics for BI requirements toward the modeling of DW and ETL process structure. This will difficult the process to define the DW and ETL process specifications accordingly. This paper adopts the Requirement Analysis Method for ETL Processes (RAMEPs) framework for the BI system approach and focused on requirement analysis method for designing the DW schemas and ETL process specifications. The RAMEPs is based on ontology and goal-driven approach in analyzing the business requirements and modeling the data integration and transformation activities. A case study of the student affair in University domain is used to illustrate how the BI approach can be implemented.

Keywords: Business Intelligence, Data Warehouse, ETL Processes, Ontology, Requirement Analysis

INTRODUCTION

Business Intelligence (BI) system using the Data Warehouse (DW) for extracting, storing, processing, and providing data to present complex and meaningful information for decision makers. These data are collected, stored, and accessed through Extract, Transform, Load (ETL) processes in order to maintain the readiness of data in a centralized DW. The data processing role is crucially important to ensure the success of the BI systems, which is dependent on the DW structure and ETL process specifications (Öykü, Mary and Anna, 2013). The design process of BI is difficult due to ambiguity of user requirement, non-standardization of DW modeling and designing methods, and lacks of tools to support the design tasks. Moreover, the design tasks need to tackle the complexity of DW and ETL process from the early to late phases of system development to ensure the user requirements are properly mapped to the DW structure for implementing the BI systems. Requirement analysis of BI processes focuses on the transformation of informal statements of user requirements into a formal expression of ETL process specifications. The informal statements are derived from the requirement of stakeholders and analyzed from the organization, decision-maker and developer perspectives

(Azman, Syazwan and Norita, 2011). The analyzing of the BI requirements from the goals of an organization, decision-maker and developer toward the detail of data transformation is important in tackling the complexity of DW and ETL process design. It is widely accepted that the early requirement analysis can significantly reduce the possibility misunderstanding of user requirements (Yu, 1995; Lamsweerde, 2009; Dalpiaz, Giorgini, & Mylopoulos, 2013). The better understanding amongst organization, decision-maker and developer, the higher are the chances of agreeing on terms and definitions used during the data transformation. Therefore, this paper utilizes Requirement Analysis Method for ETL processes (RAMEPs) method to support the implementation of the BI approach.

THE RAMEPs

The aim of RAMEPs is to facilitate the design of DW and ETL processes by analyzing requirements and producing DW schema and ETL process specifications as required by the users (Azman, Syazwan and Norita, 2011; Maté, Trujillo, & Yu, 2013). Through RAMEPs, the DW and ETL processes are modeled and designed by capturing two important facts in the DW model: i) DW schemas, and ii) data source integration and transformation activities. Therefore, the RAMEPs approach used organizational modeling to identify goals that are related to facts, and attributes of the required information, decisional modeling to identify goals for decision makers that related to facts, dimension, and measures, developer modeling to identify goals for developers that related to actions and business rules for the data transformation activities. As shown in Figure 1, these models are combined and performed as a RAMEPs method to facilitate the design of the DW and ETL process specifications.
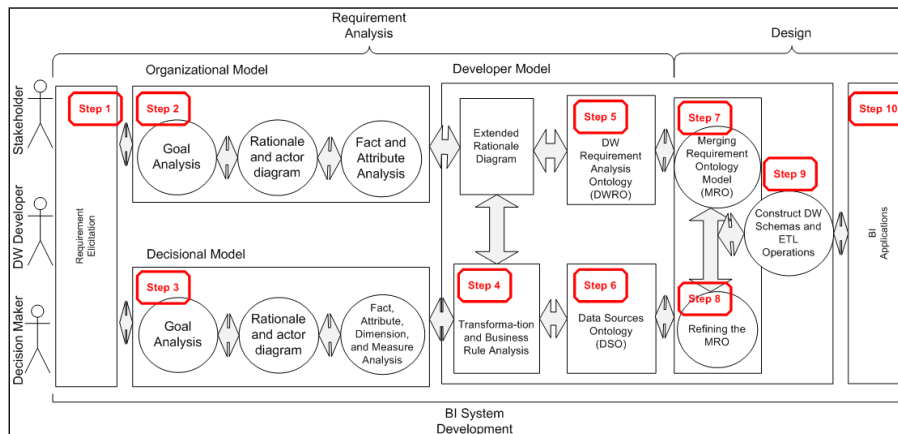

Figure 1. The RAMEPs

Based on RAMEPs model presented in Figure 1, Table 1 highlights the implementation steps of RAMEPs.

Steps	Activities	Stages of RAMEPs	Method	Output
1.	Elicit requirements.	Requirement Gathering and elicitation.	Interview, presentation, discussion.	Requirement Documents
2.	Analyze requirements on the organization perspective.	Organizational-based analysis on facts, and attributes.	Tropos – Goal-Oriented	Diagram of Organization Model
3.	Analyze requirements on the decision-maker perspective.	Decisional-based analysis of facts, dimensions, and measures.	Tropos – Goal-Oriented	Diagram of Decisional Model

4.	Analyze requirements on the developer perspective.	Data sources, Business rules, and transformation analysis.	Tropos – Goal-Oriented	Diagram of Developer Model
5.	Ontology construction on the requirement analysis glossaries	Ontology model of requirements analysis.	RDF/OWL	Ontology for Requirement Glossaries
6.	Ontology construction on the data source schemas	Ontology model of data source schemas.	RDF/OWL	Ontology for Data Sources
7.	Mapping and merging the requirements ontology with the data sources ontology.	Conceptual model of DW and ETL processes.	RDF/OWL	Merging Ontology
8.	Refine the structure of merging ontology.	Conceptual model of DW and ETL processes.	RDF/OWL	Refine Merging Ontology
9.	Constructing the required DW schemas and ETL processes specifications.	Conceptual model of DW and ETL processes.	RDF/OWL, Java and Jena 2 Framework	DW schemas and ETL Processes Specifications

THE CASE STUDY - UNIVERSITI UTARA MALAYSIA (UUM)

UUM has developed a University Management Information System (UMIS) to support the university functions as required by the users such as students, operational staff, management staff, Ministry of Higher Education (MoHE), and the public. UMIS comprises several main applications that are implemented in different databases. These applications consist of Academic Student Information System (ASIS), Graduate Academic Information System (GAIS), Personal Information System (PERSIS), Integrated Financial and Accounts System (IFAS), and others are integrated as shown in Figure 2.
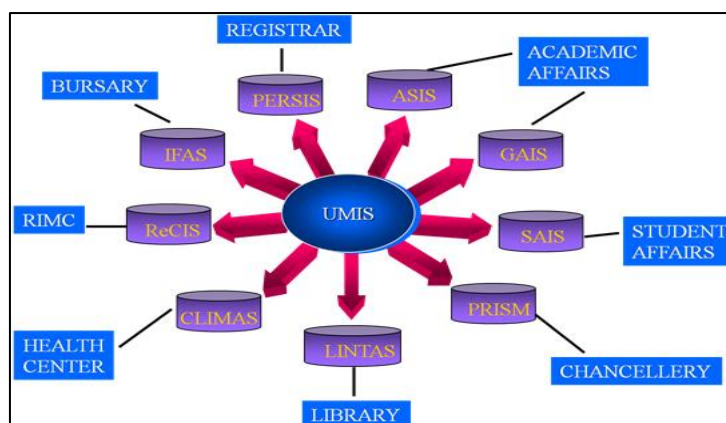

Figure 2. University Management Information System (UMIS)

This case study focuses on the BI system for producing information of the student affairs that produces from ASIS and GAIS. Originally, these systems were designed by different depart-

ments and are entirely managed by the Academic Affair Department (AAD). However, these systems were implemented in different databases, and therefore was facing the heterogeneity problems during the data integration and transformation. This research has designed the DW and ETL process according to RAMEPs approach.

Goal-Oriented Requirement Analysis

Based on the interview, the university goals are identified and details of the AAD goals are explored in supporting the university's main goals. The university goals are shown in Figure 3. To simplify the process, the case study focuses on the student affairs. The sub-goal to be the *Center of Excellence in Management Education* is relevant to the business tasks of AAD. Thus, the next task of requirement analysis is focused on this sub-goal. The scenario of student affairs that needs information from the BI system to support the goals can be described as follows:

> *"The AAD depends on the student for achieving the excellent student and depends on the lecturer for the goal of creating a culture of academic excellence. Moreover, the lecturer depends on the student for the goal of providing excellent teaching and learning"*
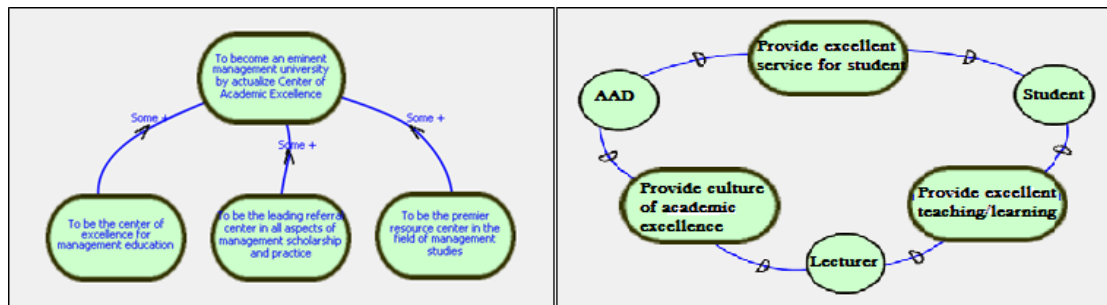

Figure 3. Goal and Actor Diagram for UUM

After the analysis is completed, the information about facts, dimensions, attributes, measures, actions, and business rules are presented in the DW requirement's diagram. The diagram for student registration and student performances is shown in Figure 4 and Figure 5 respectively by using DW-Tool (Giorgini, Rizzi, & Garzetti, 2008). These represent the final DW requirements before proceeding to the ontology model. Based on this diagram, the DW schemas (i.e., dimensions and measures), and ETL process activities (e.g., count student registered) are suggested.
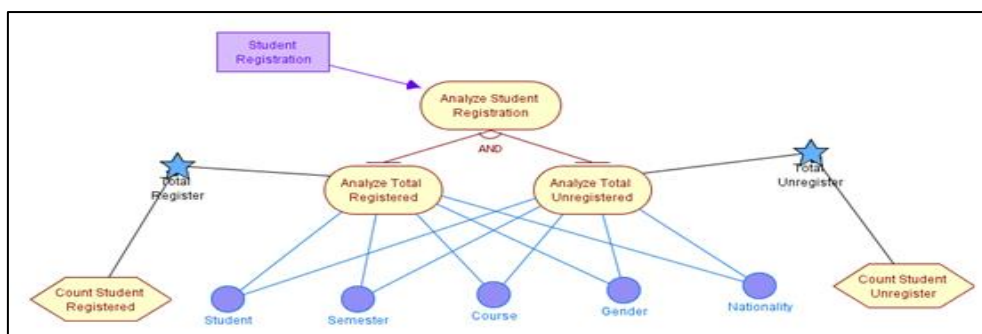

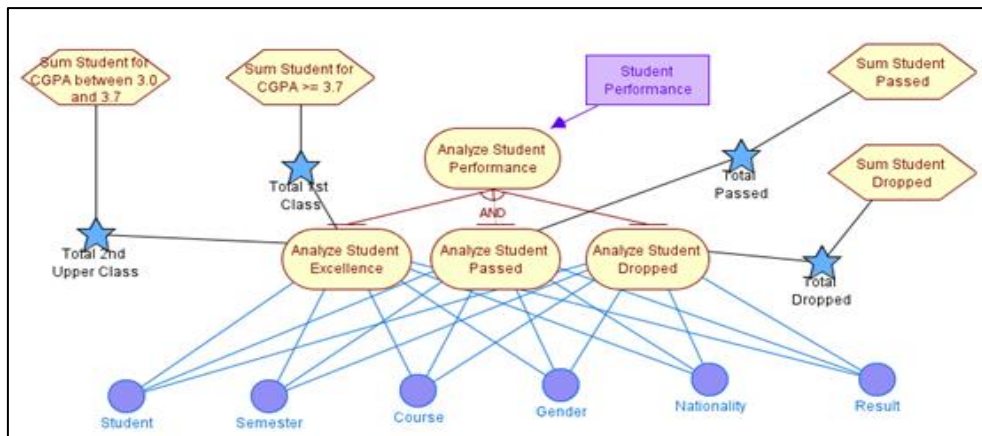Figure 4. Student Registration Goal Diagram

Figure 5. Student Performance Goal Diagram

The Analyze Student Registration and Analyze Student Performance goal diagram that defined the DW schemas is summarized in Table 2.

Table 2. DW Schemas for Student Registration

Table Type	Description/Component	
	Registration Goal Diagram	**Performance Goal Diagram**
Fact	Student Registration	Student Performances
Dimension	Student, Semester, Course, Gender, Nationality	Student, Semester, Course, Gender, Nationality, Result
Measure	Total Registered, Total Unregistered	Total 1st Class, Total 2nd Class, Total Passed, Total Dropped
Action	Count Student Registered, Count Student Unregistered	- Sum Student for CGPA between 3.0 and 3.7 - Sum Student for CGPA greater or equal to 3.7 - Sum Student Passed, Sum Student Dropped
Business Rules	A student must be Malaysian	A student must be Malaysian

Ontology Modeling

The DW components such as facts, dimensions, attributes, actions, and business rules are modeled in ontology as a conceptual design of DW and ETL processes. The ontology is constructed based on the defined model **O = (F, D, M, Br, Ac)**. Set of classes representing the concepts of the facts, dimensions, and measures, set of properties representing relationships between facts, dimensions, and measure, and set of axioms used in defining the business rules, actions, and relationship between classes are given. All these definitions are translated into the ontology model (i.e., Data Warehouse Requirement Ontology - DWRO). The ontology model includes the ontology for data sources ASIS and GAIS (i.e., Data Source Ontology - DSO).

The mapping process involves the identification of similarity and dissimilarity of concepts and their associated attributes toward the data sources. These elements are represented in the ontology structure as follows:

- The concept is represented by classes such as Student Registered, Student Performance.
- The relationship is represented by properties such as hasDimensionStudent
- Specific element in DW is represented by new classes such as SUM, COUNT
- Restriction is represented by axioms such as "A student must be Malaysian"

The ontology mapping between DWRO and DSO is shown in Table 3. These mappings should not change the semantics of user requirements as presented in DWRO.

Table 3. DWRO and DSO mapping for Student Registration

DWRO	DSO	The mapping elements (DWRO ↔ DSO)
Fact (Student Register)	-	Concept: Student Registration
Dimension (Student, Semester, Course, Gender, Nationality, Result)	Concept: Student Profile (t210student, t801studmas) Concept: Sex (t012jantina, t801jantina) Concept: Session (t005term, t005termx) Concept: Program (t006program, t808kursus) Concept: Race (t013bangsa, t801ras)	Student ↔ Student Profile Semester ↔ Session Course ↔ Program Gender ↔ Sex Nationality ↔ Race * Result is not applicable in this Fact. Thus, no mapping is established.
Measure (Total student register, Total student Unregister)	- Concept: Student Profile for status active - Concept: Student Profile for status inactive	[Total student register] ↔ Student (Active) [Total student unregister] ↔ Student (Not active)
Business Rule ("Student must be Malaysian")	Concept: Race (t013bangsa, t801ras)	[Student must be Malaysian] ↔ [Race]
Action (COUNT for Student Register, COUNT for Student Unregister, FILTER for Student must be Malaysian)	Concept: Student Profile (t210student, t801studmas), Concept: Race (t013bangsa, t801ras)	[COUNT for Student Register] ↔ [Student Profile is active] [COUNT for Student Unregister ↔ [Student Profile is inactive] [FILTER Student must be Malaysian] ↔ [Student Profile JOIN Race is Malaysian]

Table 3 presents the mapping elements of DWRO and DSO that were derived from the analysis process of user requirements and supported by the related data sources. However, to complete the entire cycle of DW schemas and ETL process design, the tasks must have actions for extract, transform and loading functionalities. These functionalities are the generic activities for extracting and loading data sources to the DW after transformation activities are completed. Based on the mapping results, new classes and properties pertaining to the merging ontology are produced. Example of these new classes is TOTAL STUDENT REGISTERED and COUNT. These new classes are defined into the merging ontology through Protégé-OWL. Moreover, the merging process is done through the ontology setting as defined in Table 4. This setting example is for Student Registration merging ontology.

Table 4. Setting for Ontology Merging of Student Registration

MAPPING LIST	ONTOLOGY SETTING
MERGE ASIS, GAIS	<u>Classes</u> Student : t210student ∪ t801studmas Gender : t012jantina ∪ t801jantina Session : t005term ∪ t005termx Course : t006program ∪ t808kursus Race : t013bangsa ∪ t801ras MergeSources: hasMergeStudent *some* Student, has-MergeGender *some* Gender … <u>Properties</u> hasMergeStudent(Domain:Student, Range:t210student, t801studmas) hasMergeGender(Domain:Gender, Range:t012jantina, t801jantina) …
FILTER Race for "Malaysian"	∃hasMalaysian ← Total_Registered, Total_Unregistred hasMalaysian *some* Total_registered hasMalaysian *some* Total_Unregistered
AGGREGATE (COUNT) for Student Registered	∀hasMeasureRegister ← Total_Registered hasMeasureRegister only Total_Registered
AGGREGATE (COUNT) for Student Unregistered	∀hasMeasureRegister ← Total_Unregistred hasMeasureRegister only Total_Unregistered

This process ends when the ontology structure is reconstructed and rechecked by using the reasoner (i.e., Pellet). The new structure of merging DWRO and DSO with new classes is known as the merged requirement ontology (MRO). In Protégé-OWL, each class and property is shown with a label, which explains the relationship between class to class, and class with properties. The MRO diagram is shown in Figure 6.
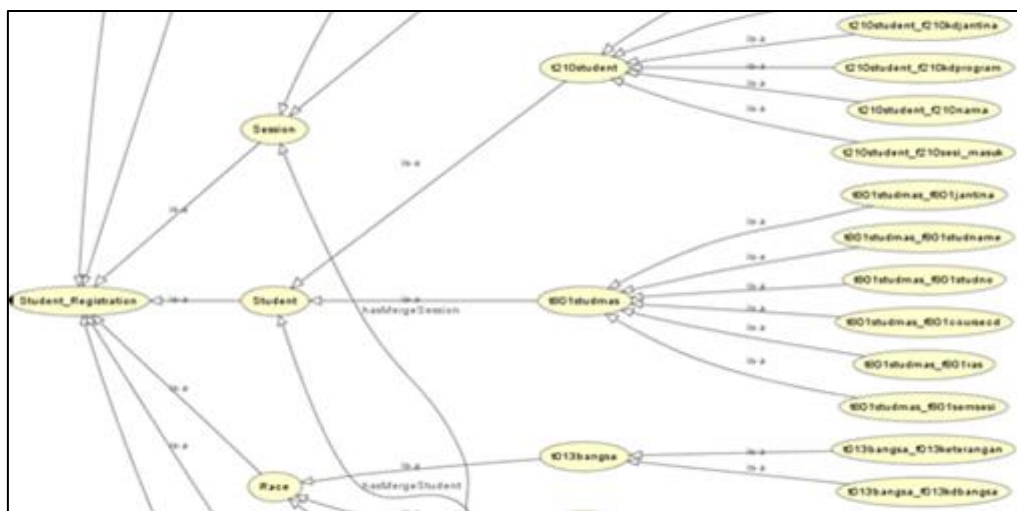

Figure 6. The MRO for Student Affairs

The DW schemas and ETL process specifications were generated based on MRO that was represented by RDF/OWL. A prototype of application for generating the DW schemas and ETL process was developed by using Java through Jena 2 Framework that runs on Eclipse

platform. The manipulation process is guided by the algorithm as proposed by Azman, Syazwan and Norita (2011). The expert reviews were conducted to clarify the strengths and weaknesses of the BI development approach. The exemplar questionnaire method is used for evaluating the approach, especially for requirement engineering approach (Cysneiros, Werneck, & Yu, 2004). A set of questionnaires together with the case study was given to seven DW developers, which three of them are from the government agencies, and the others are from the DW companies. The seven DW developers selected were qualified to assess the features of the RAMEPs, since the appropriate number of focused participants ranges from six to nine users (Nielsen, 1997; Sobreperez, 2008). Moreover, their experiences are within the ranges of three to seventeen years in developing and implementing the DW systems in various organizations.

THE RESULTS AND DISCUSSION

The results have shown that the DW schemas and ETL process specifications can be derived from the early stages of BI system development. These specifications represent data transformation for producing the information such as Total Student Registered, Total Student Unregistered, Number Student First Class, and Number Student Dropped. The ETL process specifications were used into the ETL process modeling tools (i.e., PowerDesigner version 15.3), and each of the activities was translated into the SQL statements for BI system implementation. The sequence of ETL process executions will follow the results as produced from the DW and ETL generation process. However, the order of the ETL process may not necessarily follow the sequences since the best practices still depend to the developers efforts and knowledge. The adoption of RAMEPs was helping the developers to design the DW and ETL processes until the implementation of the entire BI system. These specifications were assisted the developers to accelerate the development process through any tools that support the modeling and implementation of the BI system. Moreover, the DW design from an organization perspective by highlighting the organization's goal, alignment between DW goals and an organization's goal, goal-oriented information requirements' modelling and deriving the DW schemas and ETL process specifications are crucially important (Cravero, Mazón, & Trujillo, 2013).

CONCLUSION

The adoption of goals and ontology approach can help developers to clearly define the user requirements prior to the detailed design of the DW and ETL process in a BI systems environment. The ontology helps developers to resolve semantic heterogeneity problems during data integration and transformation activities. The RDF/OWL language is easy to maintain the DW and ETL process specifications through Protégé-OWL, although the changes in user requirements is frequently occurring. The proposed BI model has proven the DW and ETL process specifications can be derived from the early phases of BI systems development. The methodology used in analyzing the user requirements was validated by DW-Tool and Protégé-OWL successfully. Furthermore, the evaluation was implemented in university domains of case studies and reviewed by the BI experts for identifying strengths and weaknesses of the approach. Moreover, this BI approach can be used in Enterprise Application Integration (EAI) and Enterprise Information Integration (EII) environments to achieve a consolidated data for BI reporting. The semantic web-based applications for BI seem to have promising prospects to adopt this approach.

REFERENCES

Cravero Leal, A., Mazón, J. N., & Trujillo, J. (2013). A business-oriented approach to data warehouse development. Revista Ingeniería e Investigación, 33(1).

Cysneiros, L. M., Werneck, V., & Yu, E. (2004). Evaluating Methodologies: A Requirements Engineering Approach Through the Use of an Exemplar. Paper presented at the 7th Workshop on Autonomous Agents.

Dalpiaz, F., Giorgini, P., & Mylopoulos, J. (2013). Adaptive socio-technical systems: a requirements-based approach. Requirements engineering, 18(1), 1-24.

Giorgini, P., S. Rizzi, and M. Garzetti (2008). GRAnD: A Goal-Oriented Approach to Requirement Analysis in Data Warehouses. *Decision Support Systems*, 45, 4-21.

Lamsweerde, A.v. (2009). Requirements Engineering - From System Goals to UML Models to Software Specifications, John Wiley & Sons Ltd. 682.

Maté, A., Trujillo, J., & Yu, E. (2013). Aligning Data Warehouse Requirements with Business Goals.

Nielsen, J. (1997). The use and misuse of focus groups. Software, IEEE, 14(1), 94-95.

Öykü Işık, Mary C. Jones, Anna Sidorova (2013). Business intelligence success: The roles of BI capabilities and decision environments, *Information & Management*, 50(1), 13-23.

Sobreperez, P. (2008). Using plenary focus groups in information systems research : more than a collection of interviews. *Electronic Journal of Business Research Methods*, 6(2), 181-188.

Ta'a, A., M.S. Abdullah, and N.M. Norwawi (2011). Goal-Ontology ETL Processes Specification. *Journal of Information and Communication Technology (JICT)*, 15-43.

Yu, E. (1995). Modeling Strategic Relationships for Process Reengineering, in Department of Computer Science, University of Toronto.