María del Mar Gallardo (Ed.)

JCSD 2015
Actas de las
XXIII Jornadas de Concurrencia y Sistemas Distribuidos

DCTI 2015
Actas del
Doctoral Consortium en Tecnologías Informáticas

Málaga
10, 11 y 12 de junio de 2015

Jornadas organizadas por el Mobile Networks and Software Realibility Lab de la Universidad de Málaga en el Edificio de investigación Ada Byron

ISBN: 978-84-606-8792-4
Depósito legal: MA 891-2015

Enterprise Intelligence based on Ontology Metadata

Francisco Guimarães[1], Carlos Caldeira[2], Paulo Quaresma[3]
[1] Departamento Informática, Universidade de Évora, Portugal
francisco.santana.guimaraes@gmail.com
[2] Departamento Informática, Universidade de Évora, Portugal, ccaldeira@di.uevora.pt
[3] Departamento Informática, Universidade de Évora, Portugal, pq@di.uevora.pt

Abstract. Organizations define business models as part of its strategic thinking from which build performance evaluation structures focused on the effectiveness and efficiency of their goals. Usually the business model is captured in various organization representations with little interoperability between them. On the other hand, the performance are evaluated using business intelligence systems. Despite the usage of metadata in business intelligence and organization representation model system, metadata are normally not reused for the purpose of ensuring business concepts alignment. This article consolidates a vision of organizational metadata from various forms of representation of the business model, but implemented as ontology to support an organizational intelligence.

Key words: Business Model, Business Intelligence, Enterprise Architecture, Metadata, Ontology, Data Lineage, Ontology Learning, Knowledge Management.

1 Introduction

The management of organizations requires from information systems not only features to optimize the effectiveness and efficiency based on the information, but also to consider the implementation of an integrated architecture of relationships between strategy, processes, organic, computer applications and supporting technology , seen as structural assets on which organization activities take place. These assets, with is structures and business rules, are part of the business model and are caught in enterprise architecture tools, information systems models, databases and particularly in management information systems (Business Intelligence). However, there are integration and interoperability issues that inhibit a holistic view of the organization from the consolidated information independently of these models.

Despite being a concept used in business intelligence systems, metadata is a structure and concept also used in data modeling tools, modeling systems and enterprise architectures. This concept is the basis for implementation of trends in Data Governance processes transversely in organizations. On the other hand, the metadata representation forms implemented as Ontologies allow align these concepts, with clear benefits for information management but within a specific corpus of each organization activity sector and its specific language by taking advantage of the inference mechanisms on ontologies. Currently, this alignment has been the subject of research in transformation of relational databases on semantic databases or vision of entrepreneurial architectures as ontologies, but without specific investigations in the Metadata view as ontology in terms of Business Intelligence, despite some investigations in this field [1],[2],[3] and [4].

This article aims to present the advantages and feasibility of an integrated cross metadata structure, implemented with concepts of ontologies using OWL (Ontoloy Web Language) to support the concept of Semantic Web and thus allow the inference mechanism and natural language processing as solutions to Data Discovery and Data Lineage. This hypothesis solution addresses the problem of dynamic adaptation of the organizations based on the dispersed and complex information about the organization itself in various models, thus, creating an organizational intelligence.

2 State of art

Considering the problem focused on the representation of the organizations business model in organizational metadata, the state of the art covers the field of organizations, business models, forms of knowledge representation and business intelligence systems.

2.1 Organization business model

An organization is a group of people working together with a careful work devison to achieve a common purpose [5]. The organization can thus be seen as a group of people organized around a particular purpose, with technical and financial resources to achieve these objectives and with a división, way of doing of labor in the form of processes. Hence the importance of the business model for organizations. Because it describes the rationale as an organization creates, delivers and get value [6]. Osterwalder also highlights the separation between strategy and business processes, and positions the business model as a link between the two concepts. This business model is a representation of the organization and the way it does business in the market in which they operate, given a description of the commercial offer, customers, processes and resources to achieve the objectives pursued. Osterwalder states that the following activities are always present in management of business model:

• **Understanding:** Capturing the structure and business logic;

• **Review:** Observe and measure the organization based on the model created;

• **Manage:** Ensuring strategic and operational management, in various dimensions, based on the common vision structure and relationships of the components;

• **Focusing:** Action of planning and resources analysis to ensure the feasibility of the plans, allowing also the position that achieve with the resources and to correct the structure and relationship of existing resources.

This activities depend on some form o organizational resources representation in a knowledge form. That's way the relevance of knowledge representation in systems like enterprise architecture, metadata or business intelligence.

2.2 Knowdlege representation

Information systems exist in a given organizational context, and the organization's architecture is a unique structure that must adapt to a reality to be represented in a data model. As such, all organizations shape in a peculiar way the concepts passing on their information systems as a way to capture this reality [7]. These concepts correspond to the representation of an object or event that enters into the composition of a system [7]. Hence the importance of understanding these concepts with origen in the business model implemented but on the form of enterprise architectures, metadata or ontologías.

In the case of enterprise architectures, according to [8] which corresponds to the standard ISO/IEC 42010: 2007, corresponds to "The fundamental organization of a system, embodied in its components, Their relationships to each other and the environment, and the principles governing its design and evolution. ". [9] States that enterprise architecture is a set of principles, methods and models used in the design and implementation of a corporate organizational structure, business processes, information systems and infrastructure. In this case, there is the emphasis on critical components of the business model, based on a logic components and derivatives relations of the concept of architecture, which is also used in the design of business models. As such, we consider that enterprise architecture is a way of putting in the form of architecture the business model. For its implementation can be used specific government frameworks (eg DoDAF, MODAF) or generic (eg TOGAF/ArchiMate, Zachman).

In the case of metadata, this concept corresponds structured information that describes, explains and locates information resources. It is the existing application knowledge or collaborators, representing aspects of internal or external to the organization, including information about business processes, rules and data structures [10]. These definition are based on the information structure of the data, to allow manage and share information, however not focusing on the semantics of the data. Therefore, other authors report that the metadata capture the data semantic residing in various sources for integration in a corporate information system [11]. Metadata can be classified as technical and business, [10],[11]:

• Technical metadada: Data structures (eg tables, relationships, fields, value domains and formats) and transformations (eg mapping rules between tables / fields and tables origin / destination fields);

• Business metadata: Descriptions of tables, fields, rules, reports, dimensions and metrics. This information depends on knowledge in areas such as industry sectors (eg account in banks, in insurance policy may be on business activity sectors, product or branch banking in insurance). Also consider the Data Lineage metadata (life cycle and relationship between origin and use of data) and Data Usage (how and for what purpose the data is used) [11].

In the case of ontologías, this concept corresponds to the formal specification or conceptual definition of ideas, concepts, relationships and other abstractions in the context of a domain [12]. As such, it is a vocabulary for use in language field, allowing communication and reuse. The formalization becomes fundamental in computing because it allows readability by machines [12]. The concept of ontologies describe various types of artifacts, including taxonomies and metadata schemes such as those used in the standard Dublin Core Metadata Initiative (DCMI) [13]. This concept allows the explicit and formal description of concepts in a domain, properties of each concept (relations and attributes) and restrictions on the properties [14]. The basic components of an ontology are classes (concepts), relations (interaction between classes), functions (relationship where the last element of the relationship can be seen as a deterministic element in relation to previous elements), axiom (meaning and restrictions, that allow modeling always true expressions) and instances (specific items) [15]. Ontologies allow us to create a semantic metadata model based on triplets object-relation object on which we can use inference engines. However, their manual creation is time consuming and complex, being necessary to use concepts of Ontology Learning [16] to improve the capture and ontology creation.

2.3 Business Intelligence systems

Business Intelligence systems aims at capturing, understanding, analysis and transformation of data into information that enables us to analyze the organization performance, using various architectural components as shown in figure 1, [17],[18], [19] and [20].

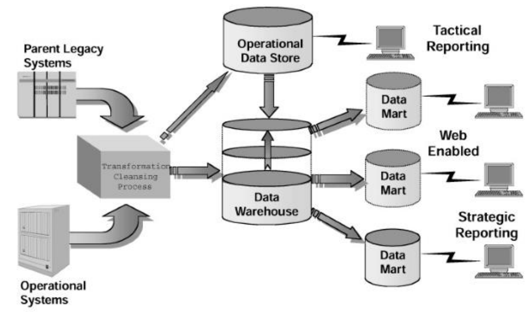

Fig. 1. Business Intelligence architectural componentes [10]

Given the various sources with different shapes, these systems store in its own database the data thus processed in what is called Operational Data Store (operational overview of sources) DataWareHouse (vision alignment concepts resolution) and Data Mart (aggregate view to exploitation by business areas such as profitability, sales, cost, efficiency, liquidity). In these databases, the concepts of metrics and dimensions, in models like star or snow flake type [19], are a type of standardized data organization, particularly in terms of Data Mart. Once the data organized in form of information ready for exploration, the analytical tools (Dashboard, Reporting, Predictive Analytics and Data Discovery) enable end-users to access them.

These systems capture and transform raw data from multiple sources, internal and external, with various formats and frequencies. This transformation process uses of ETCL tools (Extraction, Transform, and Load Cleansing) or Data Replication. One of the critical aspects of these processes are the business rules associated with the "Transform and Cleansing" where dealing with the problem of alignment of concepts and rules of transforming data into information, focusing on metrics concepts or facts (what if I want to analyze) and dimensions (where perspectives and detail/hierarchy want to analyze the metrics). The metrics represent performance measures in accordance with the events business processes, and the dimensions represent the context in which the events are processed, allowing several analytical perspectives of the same [19], with corresponds to the way of organization is structured, the business model.

The basis of transformation rules reside in models of metadata from which is possible to use Data Lineage for understanding the semantics expressed in the transformations. Data Lineage is thus a process that allows the description of where the data comes from, how they are processed and when they are processed [21]. We consider in this definition the need to show the organization structure in terms of structure and semantics representation for better understanding and exploitation in Data Discovery processes. For this reason, we focus on the metadata in Business Intelligence as a problem to find solutions via knowledge representation by ontologies as a basis for natural language processing and inference mechanisms in the field of Ontology Learning, Data Lineage and Data Discovery.

3 Problem and solution hypotesis

Information systems exist in a given organizational context, and the organization's architecture is a unique structure that must adapt to a reality to be represented in a data model [7]. Metadata allows the integrated semantic description of various data sources, providing descriptions for search, location and delivery information automatically, and ontologies have emerged as infrastructure representation of knowledge as a new approach to information engineering [22]. Based on this finding, the problem is focused research in order to consolidate the dimensions of the business model represented in several systems to enable its reuse in business intelligence as metadata, as organizational metadata. To address this issue, we focus on the solution hypothesis to implement the metadata based on ontology, to allow use of ontology learning in the creation, and inference mechanism and natural language processing on data discovery. The goal is to capture structures (tables, fields, processes, reports), relationships and attributes (characterizations of each basic concept to the specifics of subject areas and sectors of activity). The developing system to test the hypothesis is shown in figure 2.

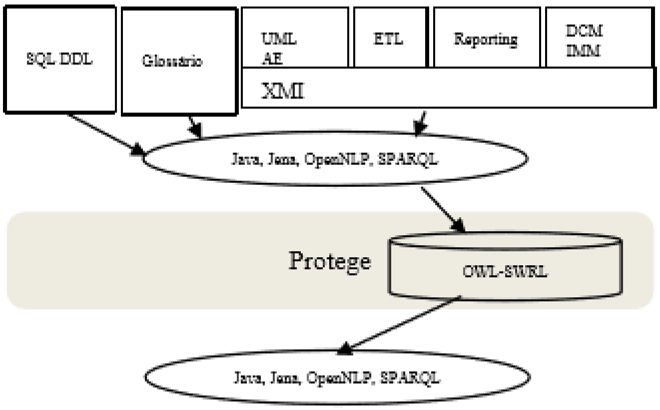

Fig. 2. System under implementation to hypotesis evaluation

The system considers a semi-automatic loading component of the ontology (using Jena, Pellet, OpenNLP and SPARQL) consolidating multiple sources to user additional definitions in Protégé tool and a model of the operating component in Java Swing for questioning in natural language processing using OpenNLP API. For its conceptualization, we analysed business intelligence systems in different activity sectors (banking, brokerage and infrastructure of water and energy) to identify the current usage of metadata, standards descriptions of concepts and type of ontology exploration of language as metadata. In this first phase, the research focused on ontology capture and the morphological and syntactic analysis of the type of expressions used in the characterization of concepts such as tables, fields and reports, and the type of expressions used in the definition of needs with the following conclusions:

• In terms of concepts, names correspond to classes, attributes adjectives and verbs are relations between classes or to indicate the class attribute list. Such is the case of "client has address, telephone number, where the customer is class and the address and telephone number are attributes, with the difference that a customer can have multiple addresses, and the address becomes a concept being implicit relationship between the concepts in the sentence;

• In terms of expressions, we standardized the structure "<verb> metric list <preposition IN> domain list <preposition BY> dimensions list <preposition FOR> list of logical conditions." Such is the case of "ANALYZE quantity, value OF Customers BY segment, region FOR year 2014 and Évora district.". Note that the implementation is in Portuguese Language.

4 Conclusions and future work

This article discusses the problem of organizations business model concepts reuse and alignment between information systems, particularly in terms of their business intelligence systems and organization representation systems.

Assuming that the business definitions are scattered in various representations to the level of database modeling tools, enterprise architectures, descriptions of specific concepts and metadata business intelligence, we designed a proposed solution that aggregates and consolidates these concepts. This cross metadata solution is supported on ontologías (OWL) Ontology Learning and uses inference mechanisms in ontology creation and natural language processing and inference mechanisms for data discovery over this metadata seen as organizational ontology. In the first phase of research we focused on the hypothesis test about ontology creation and discovery with natural language processing by interpreting common expression patterns used in business intelligence around metrics, dimensions, data domains and constraints. In the second research phase will be closed the capture rules via SWRL and analyze the implementation of Data Lineage based on inference mechanisms and compare with other cross metadata tools.

References

1. Saias, J.; Quaresma, P.; Salgueiro, P.; Santos, T.:"BINLI: An Ontology-Based Natural Language Interface for Multidimensional Data Analysis". Intelligent Information Management. 2012.
2. Chowdhury, T.; Tubb, C.:"Bridging Semantics Through Ontologies". SEMAPRO 2013: The Seventh Internacional Conference on Advances in Semantic Processing. 2013.
3. Singh, S.:"An Experiment in Software Componente Retrieval based on Metadata and Ontology Repository".International Journal of Computer Applications (0975-8887), Vol.61, N.14, January 2013. 2013.
4. Cao, L.; Zhang, C.; Liu, J.: "Ontology-Based Integration of Business Intelligence". 2006.
5. Chiavenato, I.: Comportamento Organizational – A Dinâmica do sucesso das organizações. Elsevier Editora. 2005.
6. Osterwalder, M.;Pigneur, Y.: Business Model Generation. John Wiley & Sons. 2010.
7. Caldeira, C.: "Information Ecology and Domain Definition. 6º CONTECSI – International Conference on Information System and Technology Management, Sáo Paulo. 2019.
8. IEEE Recommended Practice for Architectural Description of Software Intensive Systems. IEEE Std 1471-2000.
9. Lankhorst, M.:Enterprise Architecture at Work. Springer.2006
10. Marco, D.: "Building and Managing the Metadata Repository: A full lifecyle guide". ISBN:0471355232. 2000.
11. Kim, W.: "On Metadata Management Technology". Jorunal of Object Technology, Vol.4, No.2, March-April 2005. 2005.
12. Gruber, T.: "A Translation approach to portable ontology specification". 1993.
13. Heflin, J.: "OWL Web Ontology Language: Use cases and requirements. W3C Recommendations". 2003.
14. Noy, N.; McGuinnes, D.: "Ontology Development 101: A Guide to Creating Your First Ontology". Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880. 2001.
15. Corcho, O.;Fernandez-Lopes, M.; Gomez-Perez, A.: "Methodologies, Tools and Languages for Buidling Ontologies. Where is their meeting point? Data & Knowledge Engineering.", Data & Knowledge Engineering 46 (2003) 41–64,2003. 2003.
16. Hazman, M.; El-Beltagy, R.;. Rafea, A.: "A Survey of Ontology Learning Approaches". International Journal of Computer Applications. 2011

17. Azvine, B.;Cui, Z.;Nauck, D.: "Towards Real-Time Business Intelligence". BT Technology Journal, 23(3), 214-225, 2005. 2005.
18. Immon,W. ;Linstedt, D.: "Data Architecture: A Primer for the Data Scientist". Elsevier Kaufman. 2014.
19. Kimbal, R.; Ross, M.:"The Data Warehouse toolkit": The Definitive Guide to Dimensional Modeling", Third Edition, John Wiley & Sons, ISBN:978-1-118-53080-1. 2013.
20. Marco, D.:"Building and Managing the Metadata Repository: A full lifecyle guide", ISBN:0471355232. 2000.
21. Ikeda, R.; Widom, J.: "Data Lineage: A Survey", Technical Report, Stanford University. 2009.
22. Sicilia, M.:"Metadata, Semantics and Ontology: Providing meaning to information resources". International .Journal of Metadata, Semantics and Ontologies, Vol.1, No.1. 2006.