Opportunity Analysis for Enterprise Collaboration between Networks of SMEs

Muhammad Naeem
Decision and Information for Production Systems (DISP),
University Lumière Lyon 2, France
Muhammad.Naeem@univ-lyon2.fr

Néjib Moalla
Decision and Information for Production Systems (DISP),
University Lumière Lyon 2, France

Yacine Ouzrout
Decision and Information for Production Systems (DISP),
University Lumière Lyon 2, France

Abdelaziz Bouras
Decision and Information for Production Systems (DISP),
University Lumière Lyon 2, France
Qatar University, ICT Qatar Chair,
College of Engineering, Doha,
Qatar

Abstract

The competitive environment within the corporate have forced the small and medium enterprises (SMEs) to be more dynamic in adapting new business strategies. To achieve this objective, SMEs resort to enterprise collaboration to ooze out more business opportunities. Eventually, this culminates into a need for more useful enterprise collaboration to develop the integrated products. This aspect has accelerated the SMEs to adopt high level mechanism to move on from simple analysis to deep learning in order to realize more profit. This study is focused towards investigation of exploiting the Big Data capabilities to find out the potential opportunity lurking in the versatile nature of ever increasing organizational data. We have highlighted the issues of interoperability in the paradigm of avalanche of data, the analysis of potential opportunity as a result of enterprise collaboration leading to an added value. The analysis have been collected at surface level and then integrated into deep learning. The outcome of the study is business assets wherein the ontological modelling has been used at intermediate level.

1. Introduction

Today's business intelligence community is facing a new type of problems with deep roots in exploding nature of data in three dimensions namely Variety, Velocity and Volume (Shvachko et al., 2010). The answer to the challenge of volume lies in the analogy where the outputs from a product recommendation system dramatically undergo positive changes as the size of the data turns into big data. The solid patterns can be realized more accurately in the case of millions or billions of transactions. We take an example of a motor mechanic who can build his/her product if provided with the adequate number of relevant tools and materials. This is analogues to software development in the domain of business intelligence; but what about the scenario with teeming of versatile volume. The data scientist is

In: M. Zelm (eds.): Proceedings of the 6th Workshop on Enterprise Interoperability, Nîmes, France, 27-05-2015, published at http://ceur-ws.org

left with no option but to overlook each and every benefit of cross-utilizing these disparate and heterogeneous volume of data. Certainly, alike the raw materials and tools required to make a motor mechanic, task-specific data is best suited requirement for a data scientist. However, despite these unforeseen difficulties, integrating heterogeneous data can disseminate an interesting insight which is more attractive than exercising upon single type of data. As an example, we take data from conventional relational DBMS and subject it to business intelligence mechanism. It is likely to generate association rules, classification results, clusters, patterns and trends.

Here the question arises, which mechanism is more suitable to cope up this situation, at individual level, numerous solutions are available such as performing machine learning algorithms on structured data. Another operation is performing SQL analysis on relational data. However, these results usually are limited to explanation of inter and intra feature relationship and specification. The usefulness of trends of data generation becomes useless when the manufacturer at SME observe limited usefulness within the data being generated. This situation gives us the challenge that there must be some mechanism for deep learning. Nevertheless, it does not qualify to undermine the capabilities of established technology of data mining techniques because the path of deep learning passes through surface learning.

Big data technologies can be applied on semi-structured data by means of application of NoSQL (Shvachko et al., 2010). Products as well as services leveraging structure and semi-structured data obtained from various sources permit better organizational efficiently. It helps the data scientist to facilitate cutting edge business models endowing with deeper understanding of business needs. The business needs here confer to customer collaboration, new product design, and optimized utilization of underlying resources. From computational perspective, it leverages the business intelligence developers to innovate while uncovering the possibilities of interconnected enigmas governing an enterprise. The contribution of data and its impacts for business to an enterprise is a widely debated topic in commercial business (Naeem et al., 2014). In fact, the data is an intangible knowledge asset for any organization (Denicolai et al., 2014).

There are certain issues related to the assets impact of data. Based on this brief introduction, we shall formulate the research question; "how to define data and information assets in an enterprise" and "find out the unique characteristics associated with this data". Another aspect of data assets is the key concepts of data and quality of implicit or explicit information. More important is to address the issue of the business impact of having low-quality data and information assets. We in this study have proposed a framework. The proposed framework enables the transmission, collection, and storage of structured data in native formats. This capability of the proposed framework scales the data infrastructure to be exercised in a cost-effective manner with the increasing volume and developing new formats. The framework also brings the answer to the fundamental challenge of how to enable modern business analysis despite its complexity and diversity, for the discovery of insights in the form an opportunity analyzer.

The rest of the article consists of three sections. We shall discuss some of the most relevant research work in section 2. Advancing this study, we have proposed our model (section 3) with illustration of the results carried out (section 4). The last section is conclusion in which we have provided the overall summary of the research work in this study along with the future work.

2. Literature Review

The ever growing complexity of the enterprise data poses novel challenges in various dimensions. Obviously, enterprises are looking for innovative products or operational methodology. The fact is that large organizations have already realized this ultimate value of the data as described in the challenges in previous section. These companies such as Google, Yahoo, and Facebook etc. are already utilizing the data to provide dynamic but relevant recommendations to their users. However the question arises whether Small and Medium Enterprises (SMEs) can also exploit the data in the same way. This study gives the illustration of how is it possible for the SMEs to exploit the data to turn it into a value. While processing large amount of data, latency is also an important factor. Chelmis (2013) studied the exploitation of big data technologies for working collaboration with focus on some interesting questions including user's communication behavioural pattern dynamics and characteristics, statistical properties and complex correlations between social and topical structures. However their research focused within the internal affairs of an enterprise and did not address impact of big data for product improvement between two or more enterprises. The measurement of value of the data is tightly bound to the concept of delay i.e., latency or throughput. Luckily big data tools have permitted us to control this factor to a great length by means of putting a balance between introducing cheaper hardware and volume of data to be processed.

It is worth mentioning that the enterprise collaboration in most cases have been tackled by means of ontological modelling. The idea of exploiting ontological modelling and semantic engineering for the purpose of enterprise collaboration stem during the last decade. Some enterprise ontology models were proposed during the last decade (Oleary et al., 2010). These ontologies include ARIS (Architecture of Integrated Information Systems) by (Sheer et al., 2000); in which the enterprise ontology was consisted of twelve classes and four business views. The enterprise ontology Resources–Events–Agents (REA) introduced by Geerts et al., (2002). It bears its origin in accounting database systems with the theoretical basis in accounting measurement theory. Activity Theory Enterprise (ATE) was proposed by (Oleary et al., 2010). ATE is inspired from psychology theory providing a template based approach in capturing the context of individual activity in an enterprise. According to (Oleary et al., 2010), ATE was more suitable for more than one enterprise as well as relaxed architecture. These ontologies have been discussed in literature for organizational aspect within the enterprise including interaction between various components of the enterprise. However, we noticed that these three ontologies did not discuss the issues between two or more enterprises with different manufacturing tasks. Secondly, these ontologies were mostly static and specific to a particular nature of enterprise. We shall also investigate towards those ontology systems which target the issues of enterprise collaboration. Some good examples include SnoBase Ontology Management System (Lee et al., 2006) developed at IBM T.J. Watson Research Center. SnoBase is an industry-strength ontology management system. The strength of this system lies in the fact that it uses advanced inference approaches like semantic engine. This aspect is useful if the relationships are captured in semantic representation languages such as OWL. The SnoBase uses Fact relation to store class, property and triple in an ontology. SnoBase uses SQL triggers for the purpose of reasoning provision. However the runtime depth level of trigger cascading supported in relational database management system is limited. Another functional limitation of SnoBase is its inability to support instance reasoning. While we take this management system in enterprise collaboration model, it only addresses relational data model which limits its scope for the unstructured data.

Lin et al., (2007) proposed an ontology scheme Manufacturing System Engineering (MSE). The scheme was designed for the multi-disciplinary engineering design team for inter enterprise collaboration. The prime objective of the model stays on the introduction of a mediated ontology. The mediated ontology gives the information autonomy. Individual stakeholders don't need to understand the semantic structure of the other stakeholders. The mediated ontology serves the purpose of mediation providing the liberty of attaching to his/her own preferred language of model or ontology. However the pitfall to this technique relies in the fact that there must be a mechanism to define the mediated ontology. The manual mapping is always laborious, limited to a few ontology as well as tedious task. In this perspective the idea of mediated ontology was a naive idea. To reduce the overhead of manual mapping, semi automated features for formal mapping representation can be employed.

A design of product ontology was introduced by (Lee et al., 2009). They proposed four layered ontology architecture which serves as the foundation for the design ontology with the purpose of collaboration tasks among enterprises. It is a known fact that any commercial product is always conceived through an evolutionary style. That is why, they addressed this evolving nature of product development. The flexibility of collaboration was carried out because of its coverage to all phases of the product life cycle. They proposed the idea that each and every stakeholder is concerned with any phase of the product life cycle whether it is product design, manufacturing or supply chain management. An ontology which covers all of the aspect of the life cycle is supposed to address the requirements of all of the personnel concerned. Their architecture is aimed towards assisting in communication between user (humans) and communication among software systems. The architecture however did not address the aspects of design and quality of the software responsible for new product development.

The collaboration also comes with an issue which is related to the competition as well as access control. A knowledge access control policy (KACP) language was proposed (Chen et al., 2008). They first categorize the privileges into basic and extended access control and then proposed an ontology based access control mechanism in an enterprise. They consider three domain ontologies including product, organization and activity. One aspect was missing which was related to updating of ontologies during the integration of access strategies among enterprises.

Apart from ontology, there are some efforts made for the enterprise collaboration from other technological perspectives such as Fuzzy logic system. A Collaborative Risk Evaluator (CRE) was introduced (Wulan et al., 2012).

They formulated their technique in a web service prototype. Their technique describes well about the enterprise collaboration but scope was limited to only identification of risk analysis.

3. Proposed Framework

We in the previous sections highlighted the background of the multi sourced and variety of data which is more or less an amalgam of heterogeneous data. This amalgam of data posses two characteristics; it is multilayered and secondly it is complex; from data engineering perspective, the challenge ahead is analysis of this data. In fact, the challenge is not straight forward. If we tweak it deep inside, we come to know that this is software engineering problem. How? because not a single piece of software is able to cope up this problem. The developer need to manoeuvre from various dimensions. New functional layers are required to overcome the volume, bandwidth, and latency limitations of existing relational database solutions. The World Wide Web resources are drenched in the hype around Relational Databases, Hadoop , MapReduce and NoSQL Database systems (Özcan et al., 2014). However from literature review, the revealed gap is that there is no clarity for when any one of them needs to be preferred over the other one. Secondly, the analytical workload is also questionable (Özcan et al., 2014). Hadoop framework is known for its remarkable parallel processing capabilities. It is designed to process vast amounts of structured and semi-structured data. It has attracted the research community due to its design capability of handling versatile voluminous data due to its open source commitment. Moreover, its high aggregate bandwidth across clusters of commodity hardware is also a remarkable feature.

There is only one limitation on part of Hadoop that it is not designed for tasks requiring real time processing. Secondly, its current state is limited to those developers who are most akin to programming languages instead of interactive ad-hoc querying using a declarative language such as SQL.

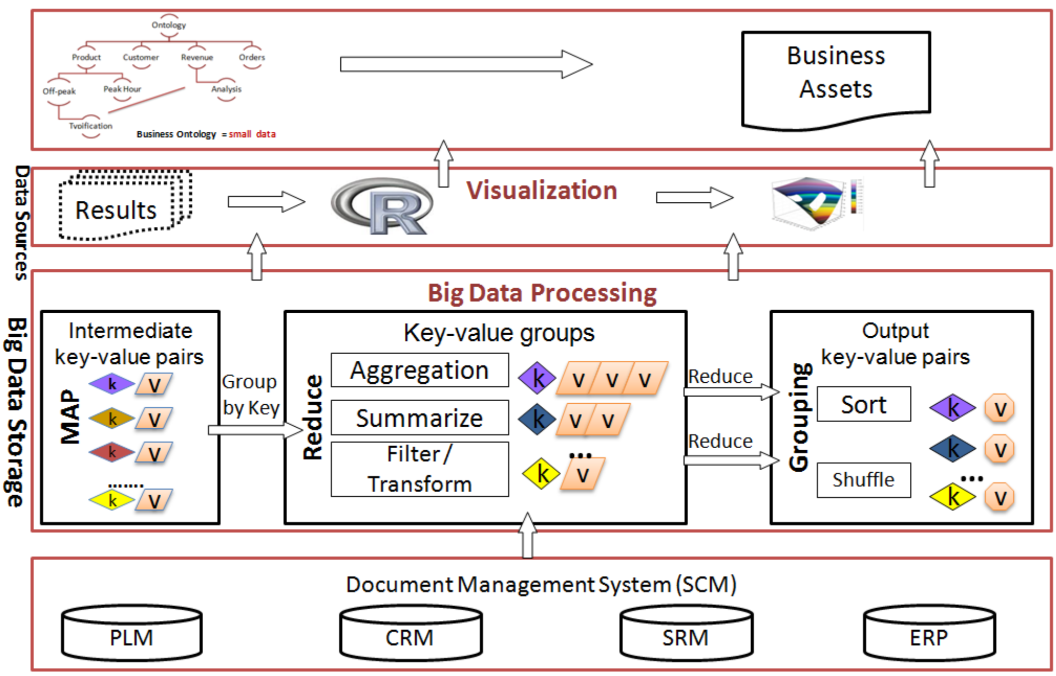

Figure 1: Functional Layer of the proposed framework

This functional framework as shown in the figure 1 is based on the axiom that the enterprises be viewed as repositories of data and knowledge highlighting the role of intangible resources. In pursuit of these objectives, the goal oriented intelligent components have been proposed for improving the mobility mechanisms of acquisition phase of the data handling mechanism. It is difficult to produce a globalized standard while achieving the goal of talking inter-enterprise collaboration (Lin et al., 2007). Figure 1 is showing four essential functional layers to achieve the results from multi-sourced data. In the first layer, we have various components which are producing the data in different shape and formats. This layer itself composed of three independent components acting as data producer.

This layer contains structured data in different formats. In fact, this component is a part of "success story", in which the wealth of corresponding and relevant data is placed using Digital Preservation model (Naeem et al., 2014). We have given this layer the name of Document Management System (D.M.S). The data in this layer is qualified from quality procedures, its Meta data in shape of user contents, quality refinements and the documents produced as a result of various types of inter-enterprise as well as intra-enterprise communication and interaction. It receives the data from the pools of the data sources producing as a result of variety of business process. These business processes may fall under Supplier Relation Management (SRM), Enterprise Resource Planning (ERP), Product Life Cycle (PLM) and Customer Relationship Management (CRM). The sequence flow of the data as shown by the figures points out that the data from D.P component is passed on to the Big Data Processing (BDP). At this point, all of this data is subjected to the acquisition phase of the Big Data technologies. The third layer is related with visualization of results. But these visual results are not sufficient. We need to translate them in to textual format. This layer also integrates the results into a higher level for the purpose of deep learning. The last layer is responsible for the ontological modelling. The ontological model is composed of an integrated ontology for better perception of business assets. In the next section, we shall discuss the practical flow of data with input and out of each layer.

4. Results and Discussion

We performed experiments on dataset provided by a local enterprise namely APR. The company deals in manufacturing of plastic products. The experiment was performed on Oracle Big Data Light Server. The company is in the business of producing numerous types of plastic products with variable parameters on demand. The company receives the quotations, negotiate the price, parameters of the product and then proceed or reject the proposal of order. Currently, the company has accumulated a large volume of datasets as a result of successful completion of massive number of orders. Our goal is to find out any opportunity lurking in structured and semi-structured enterprise data using big data capabilities. We are not restricting it to mere graphical representation of the results but the purpose is to formulate the results in form of publishable business assets. We have obtained results at different levels. The first level is comprised of the results obtained from basic SQL queries using Hive and Pig over Hadoop infrastructure. The results are on basic level. Certainly, a higher level of integration is required. By integrating these basic units of analysis, more useful patterns can be obtained. Table 1 is showing one of the higher level integrated result obtained. We obtained series of such results. These results are the integration of the results addressing the following top level queries such as: The enterprise has three types of customers placed in low privileged to high privileged category. Each category receives a specific range of discounts. What the data analysis reveal about the re classifications of its customers in these three categories? Another tangential query being addressed by this analysis addresses the question for “whether the enterprise should review about the number of its categories? Which specific business-deals generate more revenue for the enterprise? Every enterprise is always interested in formulation of the list of the customers which have relatively higher number of abandoned orders. Nevertheless, such basic inquiry leads to dig out the underlying reasons of unacceptable size of abandoned cart. Moreover, the company is also interested in churn out analysis; we carried out the analysis to find out the reasons why some valuable customers never return. Such criterion is marked by setting an upper value of threshold of profit giving customers. These type of queries stays at the primary level. If we move a step ahead, then a complex analysis can entitle our enterprise with the capability of drawing a coupon for some of its customers for certain products. This aspect gives the opportunity analysis in the domain of marketing. The enterprise can attract some of its customers based on previous sales record in two dimensions; customer wise and product wise interpretation with the added parameters of the enterprise capability to produce those targetable products on appreciable marginal profits.

Table 1: Integrated analysis of structured data

Product	Format	Mode	Charge	Dimension 1	Dimension 2	Dimension 3	Color	Quantity	Last Production
Plaque	CONSO GRANULE COULEE PU DECOUPE	Grainé Médical Moulé	Poreux OIL Antistatique Diffusant HI	12.7 14 16 45 70 80 110 140 180 300	55 - 70	260 300 310 325 330	Beige Fumé Bronze Gris Gris Bleu Ivoire Jaune Incolore	9476	May - 2014
	PETG		Prismatique						
	FABRIQUES PETG NEGOCE		Lubrifiant						
	NEGOCE		OIL						
	GRANULE	Moulé Expansé	Antistatique						
	COULEE PU		Diffusant						
	DECOUPE	Moulé	HI						

The next phase is the ontological modelling. In this phase we have two approaches. The first approach is to create a pool of ontology where each of the ontology is modelled out of a single set of analysis. The second approach is to develop a single integrated ontology. Although the later approach is complicated when there are massive number of concepts and their relationships, however a single integrated approach has always been appreciated. (Ding et al., 2002); Gene Ontology Consortium, 2015). The figure 2 is showing a part of the ontology model based on the series of analysis. The purpose of this ontology is to provide enterprise personnel with a simple and common interpretation of the business process data. This is obtained by identifying the objects (things) which are expressed in a graphical representation of the business process outcome along with its related activities, methods and techniques in spite of the philosophical assumptions. The ontological modeling justification lies in the underlying benefits achieved from the development of these general graphical models.

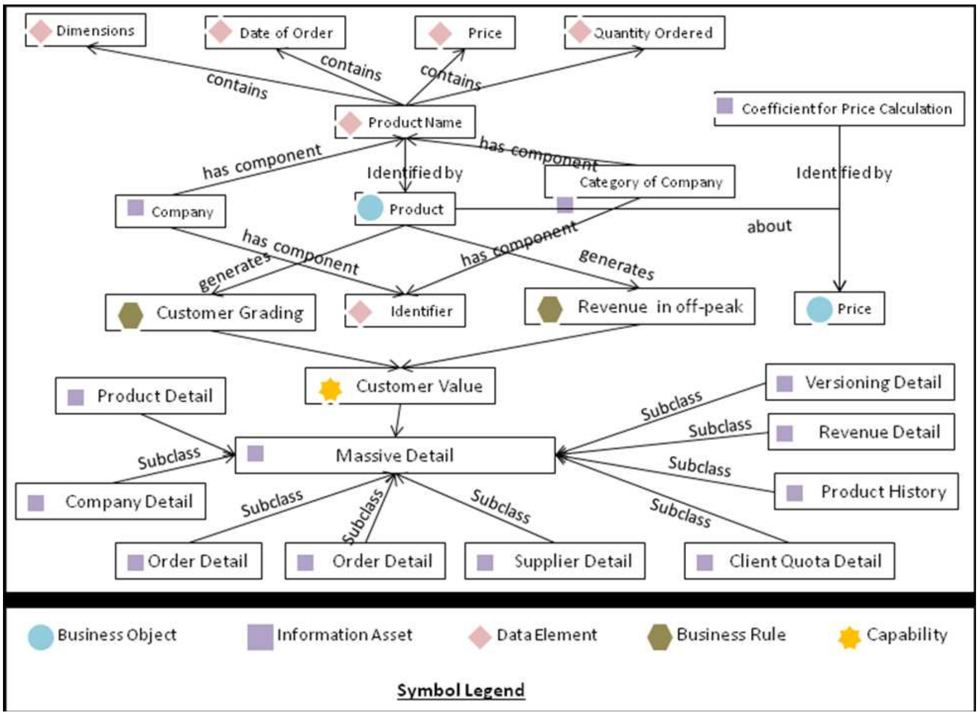

Figure 2: Relationships among Information Assets, Data Elements, and Business Objects

- First, it can help the enterprise personnel to develop visual roadmaps of the business process, enterprise overall capability to be undertaken to achieve the desired outcome. Thus, it can help the decision makers for the identification of the potential advantages, possible limitations and unforeseen bottlenecks that may arise in the anticipated roadmap based on the paradigms underpinning the business activity.
- Second, it can enhance the communication between various actors at SME such as technical persons, marketing, financial employees as well as decision makers because the models are a representation of the business's plans to be used as the base for discussion.
- Third, the graphical models in shape of ontology can help the target audience of the proposed business activity to enhance their deep understanding of the way
the business process was performed and the different decisions that were made during the formulation of the process.
-Fourth, accumulation of a reasonable number of models depicting instances of business activities can be used to categorize business activity with similar characteristics leading to identify potential opportunities within data.

The ontology is modelled by means of identifying the elements with their roles. There are five types of elements in the ontology including business object, information asset, data element, business rule and the capability. Their interaction is also defined in the data properties as well as object properties. The purpose of the ontological model is to publish the business assets which are holding the added value for the enterprise with massive amount of data. The explanation of the added value is the ultimate goal of this study. This added value serves the purpose of identification of opportunity analysis using big data technologies followed by ontological modelling. We can illustrate two examples of business assets:

1. The enterprise has the capability to produce tube plastic material within the format of Graphite, Gravage, inhouse Injection but with some specific set of modes including pressing, stabilisation while its dimension, weight and other criterions are also in stipulated range. Moreover, the company has experience of only two years and have completed more than 25000 orders without any major complaint.

2. The enterprise has the capability to produce certain products (plaques and polyamide plastic) in certain range. Our analysis has illustrated that there are 17 customers who have not ordered for these components, but they are highly anticipated customers based on big data recommendation analysis followed by ontological models.

The proposed model has been demonstrated with a lot of other similar business assets.

5. Conclusions

Amidst today's tsunami of exponential growth in applications complexity and versatility along with ever growing data, the domain of business intelligence is earning a remarkable progress towards the enterprise collaboration. In the field of enterprise collaboration, limited work has been carried out with focus on interpretation of added value by means of opportunity analysis on versatile data. We in this study have proposed a framework to investigate the added value in the masses of data. The framework has shown that the Big Data technologies can be tailored to exploit the ever-accumulated semi-structured data. The challenge was in two dimensions. The volume of the data is the aspect for which numerous typical analysis tools usually fails. The second challenge is related to unstructured nature of the dataset. The Big Data technologies have proved their worth. However our contribution is not limited to utilization of the Big Data technologies. The technical challenge was to turn the data into a stream of opportunity. Any enterprise can make a decision on the basis of this stream of opportunities for in-time collaboration with other enterprise. Moreover, we have modelled the analysis into an ontology. We then decompose the ontology into business assets serving the purpose of explanation of opportunity analysis.

References

Chelmis C., "Complex modeling and analysis of workplace collaboration data", Collaboration Technologies and Systems (CTS), 2013 International Conference on. IEEE, 2013, pp. 576-579.

Chen T.-Y., "Knowledge sharing in virtual enterprises via an ontology-based access control approach", Computers in Industry", vol. 59 no. 5, 2008, p. 502-519

Denicolai S., Zucchella A., Strange R., "Knowledge assets and firm international performance", International Business Review, vol. 23, no. 1, 2014, p. 55-62.

Ding Y., Foo S., "Ontology research and development, Part 2 - A review of ontology mapping and evolving", Journal of Information Science, vol. 28, no. 5, 2002, p. 375-388.

Gene Ontology Consortium, 2015, Gene Ontology Consortium: going forward, Nucleic Acids Research 43, no. D1, D1049-D1056.

Geerts G. L., McCarthy W. E.,. "An ontological analysis of the economic primitives of the extended-REA enterprise information architecture", International Journal of Accounting Information Systems, vol. 3, no 1, 2002, p. 1-16

Lee J., Chae H., Kim C.-H., Kim K., "Design of product ontology architecture for collaborative enterprises", Expert Systems with Applications , vol. 36, no. 2, 2009, p. 2300-2309.

Lee J., Goodwin R., "Ontology management for large-scale enterprise systems", Electronic Commerce Research and Applications, vol. 5, no. 1, 2006, p. 2-15

Lin H. K., Harding J. A., "A manufacturing system engineering ontology model on the semantic web for inter-enterprise collaboration", Computers in Industry, vol. 58, no. 5, 2007, p. 428-437.

Naeem M., Moalla N., Ouzrout Y., Bouaras A. "An ontology based digital preservation system for enterprise collaboration", Computer Systems and Applications (AICCSA), 2014 IEEE/ACS 11th International Conference on, November 2014, p. 691-698

O'Leary D. E., "Enterprise ontologies: Review and an activity theory approach", International Journal of Accounting Information Systems, vol. 11, no. 4, 2010, p. 336-352.

Özcan F., Tatbul N., Abadi D. J., Kornacker M., Mohan C., Ramasamy K., Wiener J. "Are we experiencing a big data bubble?", Proceedings of the 2014 ACM SIGMOD international conference on Management of data, June 2014, p. 1407-1408

Shvachko K., Kuang H., Radia S., Chansler R. "The hadoop distributed file system", Mass Storage Systems and Technologies (MSST), 2010 IEEE 26th Symposium on, May, 2010, p. 1-10).

Scheer A.-W., Nttgens M., "ARIS architecture and reference models for business process management", Springer., 2000

Wulan M., Petrovic D., "A fuzzy logic based system for risk analysis and evaluation within enterprise collaborations", Computers in Industry, vol. 63, no 8, 2012, p. 739-748