Applying ontologies to data integration systems for bank credit risk management

Jalil ElHassouni[1*], Mehdi Bazzi[2], Abderrahim El Qadi[3], Mohamed El Haziti[1]

1LRIT-CNRST (URAC'29), Mohammed V University- Faculty of Science of Rabat, Morocco

2SPDT, Hassan II University – Faculty of Science Ain Chock Casablanca, Morocco

3LASTIMI, High School of Technology in Salé, Mohammed V University in Rabat, Morocco

* Jalil.elhassouni@gmail.com

Abstract
This paper proposes an ontological integration model for credit risk management. It is based on three ontologies: one is global and two are local. The global ontology describes credit risk management process and the local ontologies describe credit granting process and present the concepts necessary for the monitoring of credit system. This paper also presents the technique used for matching global ontology with local ontologies.

keywords
credit risk management; data integration; ontologies alignment

INTRODUCTION

The integration of data has become increasingly important in many areas. It provides decision makers and analysts with uniform access to distributed and heterogeneous sources of data (relational databases, XML documents, semantic databases...).

However, the data integration processes across multiple sources constitute a complicated task because of the semantic heterogeneity encountered at the different levels: metadata, data schemas and data values. This issue is at the heart of many research areas, and this work falls within mapping schemas for bank credit risk management. It takes into consideration the requests of decision-makers and the indicators used in the banking industry to measure credit risks.

The main step in the risk management process is shown in Figure 1. Based on this architecture, we propose a general ontology that models the credit risk management process and two specific ontologies. The first models the process of credit allocation to clients and the second displays the concepts necessary for the monitoring of a credit system. Subsequently, we arranged the alignment of specific ontologies with the general ontology to facilitate semantic interoperation between applications based on these ontologies.
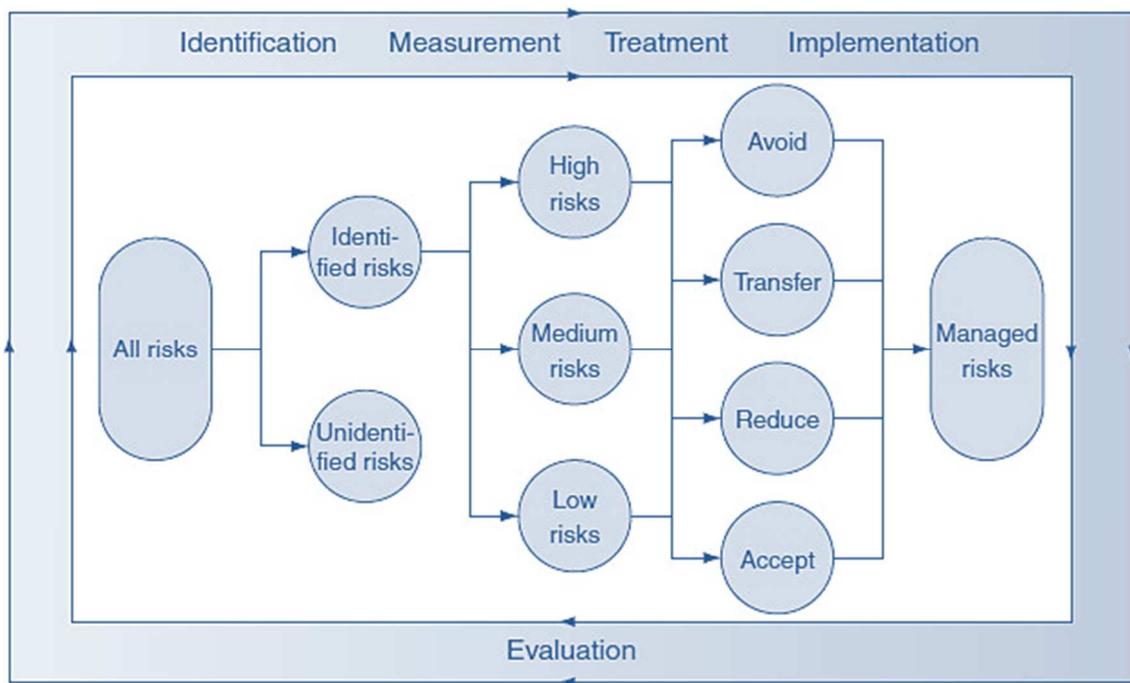
Figure 1. Different steps of a risk management process.

Figure 1 displays the different steps of risk management process:
- Identifying and quantifying risks in order to measure their impact.
- Treating risks (according to the impact and the strategy of the bank) by transferring, reducing or accepting them.
- Once the risk management strategy is defined, it is ready for implemented. Banker, statistical models and IT infrastructure assess the risk of existing and new investments. The whole cycle is continuously evaluated to detect new risks and improve the existing systems.

The remainder of this paper is organized as follows: section one provides a literature review and section two explains the process of creating ontologies in the field of bank credit risk. The last section concludes with discussing future work.

I RELATED WORK

In recent decades, several studies have been conducted in the field of data integration. Two major approaches have been set up to design data integration systems [Hacid and Reynaud, 2004]:

- The mediator approach where the exploitation of the data is based on the virtual views.
- The Data Warehouse approach (DW) where the data are duplicated in the same global schema.

In the mediator approach, the integration of data follows two steps:

- The integration system generates from a user request as many sub-queries as data sources request.
- The integration system constructs the final response from the result of each sub-query posed to each data source and transmits it to the user. This approach is well suited where information changes quickly; i.e., when the user requests are not predictable and when the numbers of data and data sources are high.

In the Data Warehouse approach, data integration is also carried out in two stages:

- The integration system merges and stores different data sources into a single data warehouse.
- The user's request searches for the data in DW without having to access the original data sources. This approach is suitable when the information is summarized and annotated or when the user requests can be predefined.

Variuos ways of applying the above-stated approaches are used different researches. [Calvanese et al, 2001] proposes a method based on a global ontology as an intermediate model between DW and source schemas. [Lamolle and Zerdazi, 2005] provide a mediation model based on XML. [Bakhtouchi et al, 2009] utilize a hybrid framework to improve the performance of queries in heterogeneous and autonomous data integration systems. [Dean and Ghemawat, 2004] and [Yuan et al, 2010] develop an approach that is based on MapReduce. [Mate et al, 2015] use an ontological approach to describe and organize the medical concepts of both data and target sources. [Krisnadhi et al, 2015] put forword an approach which combines linked data publishing and modular ontology engineering based on ontology design patterns.

However, some of these applications are limited to integrating only one source such as [Mate et al, 2015], while others such as [Krisnadhi et al, 2015] and [Berkani et al, 2012] integrate multiple sources. Several types of data thus are integrated in different ways. The approaches proposed in in [Abadie, 2012], [Zhao et al, 2008] and [Aji et al, 2013] manipulate spatial data, whereas other propositions manipulate alphanumeric data as presented in [Yuan et al, 2010], linked data in [Krisnadhi et al, 2015] or ontologies-based data as presented in [Berkani et al, 2012].

All the approaches presented above use manual or semi-automatic solutions requiring more time and domain experts' intervention. The approach provided by [Mate et al, 2015] does not support mapping at hierarchical levels between the target and source ontologies. The target ontology is too cumbersome to use and intricate. Hence, when a new data repository is added, a complicated adjustment of the ontology may become necessary. In [Krisnadhi et al, 2015] the absence of data reasoning and alignment between data repositories may provoke the incoherence and reducndacy of the target data.

II ONTOLOGIES BASED DATA INTEGRATION APPROACH

Ontology is a formal, explicit specification of a shared conceptualization in terms of concepts (i.e., classes), properties and relations [Gruber, 1995]. The term "formal" means the ontologies should be expressed in a formal and semantic language (RDF, OWL, PLIB...) allowing automatic reasoning. "Explicit" signifies ontology's acceptance by all expert community members and "Conceptualization" refers to an abstract, simplified view of the world that we wish to represent for some purposes.

According to [Gagnon, 2007] the "ontologies' integration" carries at least three different meanings (and uses) of ontology integration depending on the situation:

- Integration: building a new ontology from other existing (sub) ontologies.
- Merging: building a new ontology by merging existing ontologies i.e.: building a single ontology that unifies them all.
- Usage in applications: building an application ontology using one or more ontologies. Existing ontologies are adapted if necessary.

3

2.1. Credit allocating to customers (Credit Scoring)

Figure 2 displas the architecture of the credit scoring system. This process can divided into two phases. The first phase shows the building of the statistical model, which is characterized by a set of steps like data gathering activities, initial analysis, model development and implementation. The second phase demonstrates the use of the statistical model to make credit-related decision such as fixing the cut-off, setting credit limits, pricing and storing the information.
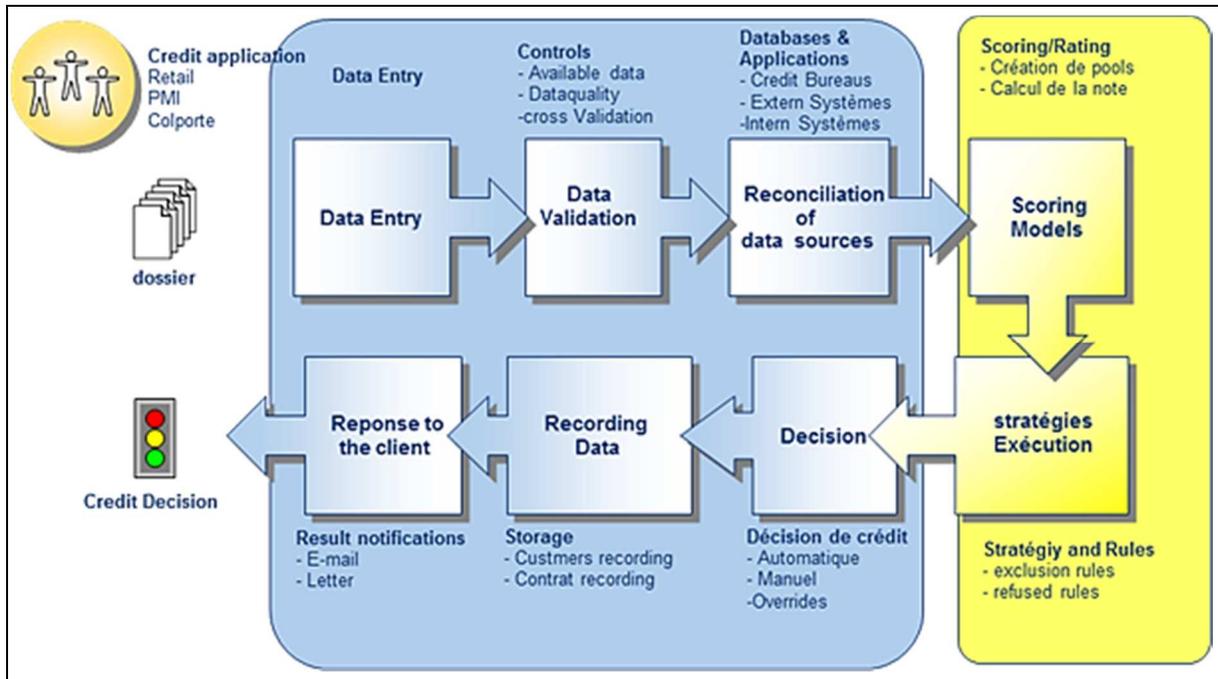

Figure 2. Process of allocating credit to customers (Credit Scoring).

Based on the risk management process presented in Figure 1, we propose a conceptual model Figure 3, where we widely use a set of variables to build credit scoring system. Some variables are related to the credit itself (Credit Requirements), while others are related to the financial profile (Financial Data) and to the personal profile (Socio Demographic) of the borrower. Other variables, however, are related to the final decision and to theborrower's rating (Decision Data). We modeled the credit allocation process on the form of ontology given in Figure 4.
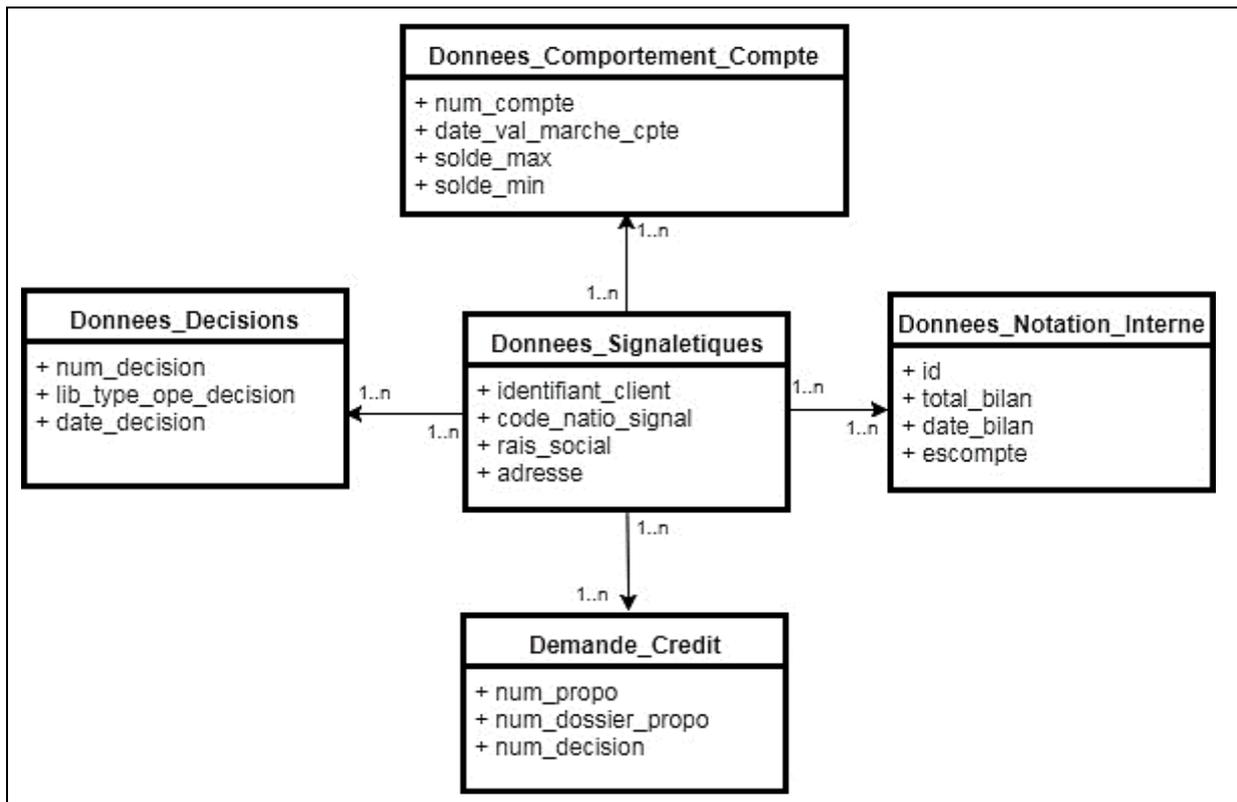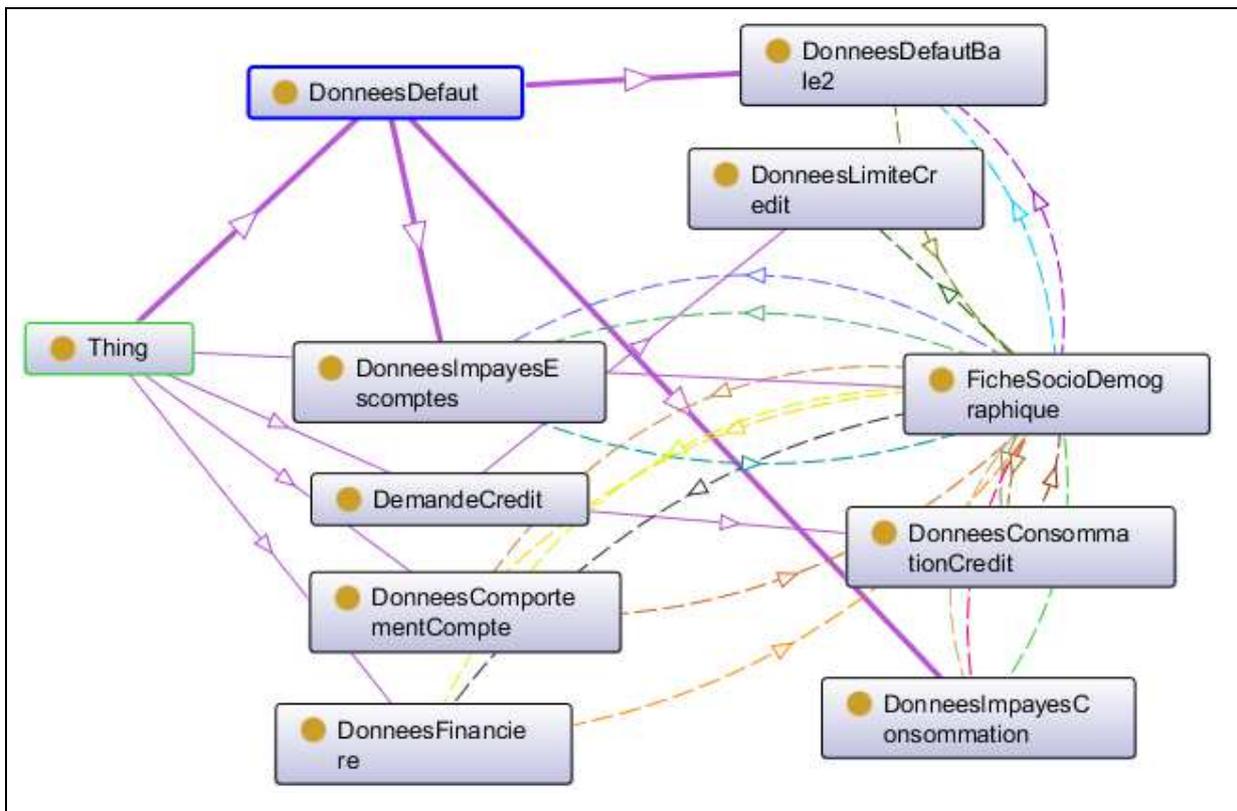
Figure 3.Conceptual model of credit risk management.

Figure 4. The ontology of the credit allocation process - partial view.

2.2. Ontology for credit monitoring

Credit monitoring is related to 'post-credit decision' collection, analysis of information, and most importantly the identification of early warning signs. Its also entails ensuring the appropriate decisions are made with regard to changes in customers' credit risk.

During the credit period, economic variations may cause some changes that have an impact on credit risk. Banks have to monitor credit exposures periodically and continuously so that changes are detected on time. The Figure 5 illustrates the main steps in credit monitoring.
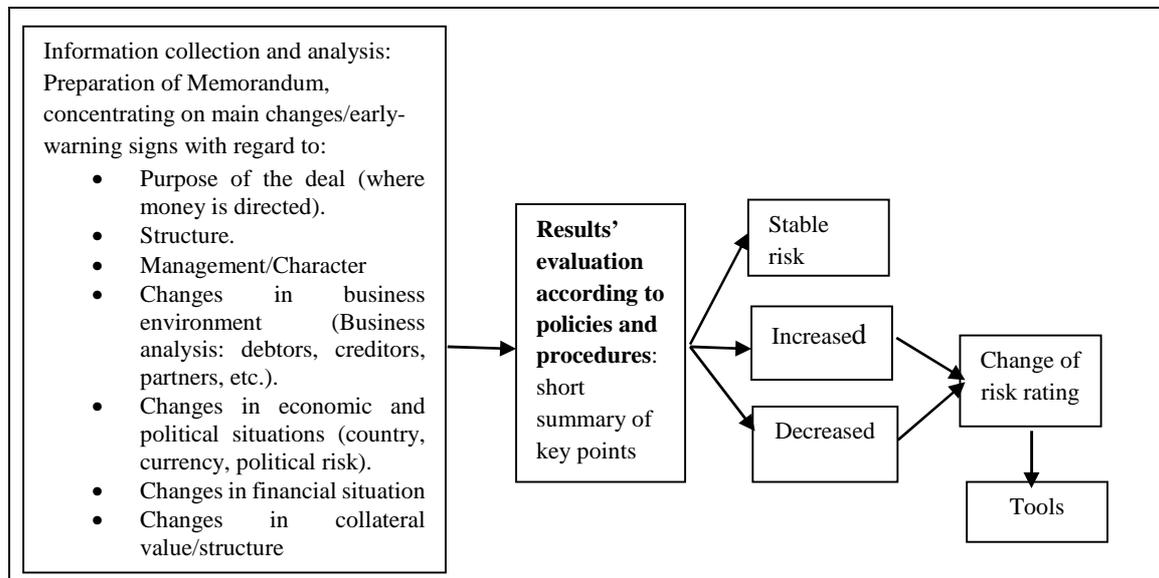

Figure 5.Main steps in credit monitoring.

2.3. Global process of risk management

Having thus defined credit allocation ontology and the credit monitoring ontology, we will build the global ontology, which will provide the user with a unified view, global schema, of data sources. In the modeling part of the global ontology, we organized several meetings and brainstorming sessions with the banking experts resulting in the development of the version described in Figure 6.
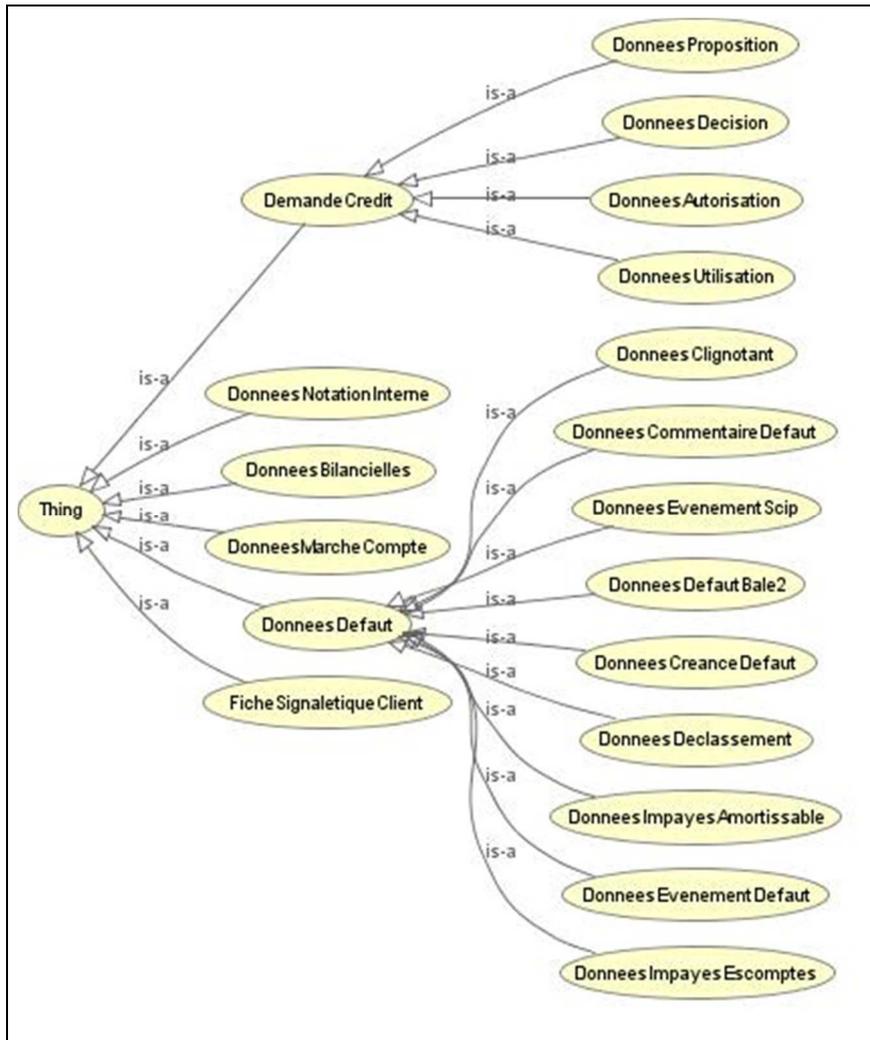
Figure 6. Monitoring of credit process ontology - partial view.

2.4. Ontologies alignment

In the last decade, several studies were released in the schema matching area, including ontology alignment. This ontology alignment is an important functionality in many applications such as data exchange, schema evolution, data integration etc. It is the basis of linking information, e.g., from heterogeneous sources into a common model that can be queried and reasoned upon. Subsequently, we have aligned the ontology of credit allocation with the global ontology in Figure 7 using LogMap Ontology Matching Tool [Jiménez-Ruiz and Grau, 2011].

```
http://www.co-ode.org/ontologies/ont.owl#DemandeCredit|http://www.co-ode.org/ontologies/ont.owl#DemandeCredit|=|1.02|CLS
http://www.co-ode.org/ontologies/ont.owl#DonneesDefaut|http://www.co-ode.org/ontologies/ont.owl#DonneesDefaut|=|0.95|CLS
http://www.co-ode.org/ontologies/ont.owl#hasFonProBilan|http://www.co-ode.org/ontologies/ont.owl#hasFonProBilan|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasMntDepasUtil|http://www.co-ode.org/ontologies/ont.owl#hasMntDepasUtil|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasAchatsBilan|http://www.co-ode.org/ontologies/ont.owl#hasAchatsBilan|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasCumulUtil|http://www.co-ode.org/ontologies/ont.owl#hasCumulUtil|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasEffComSignal|http://www.co-ode.org/ontologies/ont.owl#hasEffComSignal|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasSldmoyCrAng|http://www.co-ode.org/ontologies/ont.owl#hasSldmoyCrAng|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasTel1Signal|http://www.co-ode.org/ontologies/ont.owl#hasTel1Signal|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasAnnBilBilan|http://www.co-ode.org/ontologies/ont.owl#hasAnnBilBilan|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasMvtNetDbAn|http://www.co-ode.org/ontologies/ont.owl#hasMvtNetDbAn|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasMvtNetCrAn|http://www.co-ode.org/ontologies/ont.owl#hasMvtNetCrAn|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasProvDefaut|http://www.co-ode.org/ontologies/ont.owl#hasProvDefaut|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasAmortBilan|http://www.co-ode.org/ontologies/ont.owl#hasAmortBilan|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasSiglSignal|http://www.co-ode.org/ontologies/ont.owl#hasSiglSignal|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasSoldMoy|http://www.co-ode.org/ontologies/ont.owl#hasSoldMoy|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasMvtNetCrAn2|http://www.co-ode.org/ontologies/ont.owl#hasMvtNetCrAn2|=|1.0|DPROP
```

Figure 7.The alignment file of credit allocation ontology with the global ontology in text format - partial view.


```
    // Classes
    //
    ////////////////////////////////////////////////////////////////////////
    |
-->
<!--
 http://www.co-ode.org/ontologies/ont.owl#DemandeCredit
-->
<rdf:Description rdf:about="http://www.co-ode.org/ontologies/ont.owl#DemandeCredit"/>
<!--
 http://www.co-ode.org/ontologies/ont.owl#DonneesBilancielles
-->
<rdf:Description rdf:about="http://www.co-ode.org/ontologies/ont.owl#DonneesBilancielles"/>
<!--
 http://www.co-ode.org/ontologies/ont.owl#DonneesDecision
-->
<rdf:Description rdf:about="http://www.co-ode.org/ontologies/ont.owl#DonneesDecision"/>
<!--
 http://www.co-ode.org/ontologies/ont.owl#DonneesMarcheCompte
-->
<rdf:Description rdf:about="http://www.co-ode.org/ontologies/ont.owl#DonneesMarcheCompte"/>
<!--
 http://www.co-ode.org/ontologies/ont.owl#DonneesNotationInterne
-->
<rdf:Description rdf:about="http://www.co-ode.org/ontologies/ont.owl#DonneesNotationInterne"/>
<!--
 http://www.co-ode.org/ontologies/ont.owl#FicheSignaletiqueClient
-->
<rdf:Description rdf:about="http://www.co-ode.org/ontologies/ont.owl#FicheSignaletiqueClient"/>
</rdf:RDF>
```

Figure 8.The alignment file of credit monitoring ontology with the global ontology in text format - partial view.

Figures 7 and 8 show the mapping generated by LogMap algorithm. They illustrate the alignment of classes and data properties, which are semantically equivalent in both onthologies, and the degree of similarity between them.

Ontologies are created to follow up credits of various customers of the bank and allow for several analyses including:

- Analysis of the stability of elaborated models which requires reviewing the scoring models developed by banks each year.
- Roll rate analysis: The percentage of credit customers who become increasingly delinquent on their accounts. Roll rate refers to the percentage of customers who "roll" from 30-day-late to 60-day-late categories, or from 60-day-late to 90-day-late categories, and so on.

Conclusion

In this paper, we presented our model of creating ontologies to facilitate the integration of Credit Scoring data. We proposed a general ontology that models the credit risk management process, and two specific ontologies that model the allocation and monitoring of credit. The result of the alignment shows the entities (classes and data properties) that are semantically equivalent. In our future work we plan to model the tracking credits taking into account the requests of decision makers in the process of integration. We also plan to develop a hybrid model for credit risk management based on ontologies and MapReduce approach.

References

M.S Hacid & Reynaud. L'intégration de sources de données. La revue Information – Interaction – intelligence – (R 13), 2004.

D. Calvanese, G. De Giacomo, M. Lenzerini, D. Nardi, and R. Rosati. Data integration in data warehousing. Int. J. Coop. Inf. Syst. 2001:10(3):237–271.

M. Lamolle and A. Zerdazi. Intégration de Bases de données hétérogènes par une modélisation conceptuelle XML. COSI. 2005:5:216–227.

A. Bakhtouchi, L. Bellatreche and A. Balla. Materializing attributes annotation: A hybrid approach for databases integration. STIC 2009.

J. Dean and S. Ghemawat. MapReduce: Simplied Data Processing on Large Clusters. OSDI 2004.

Y. Yuan, Y. Wu, X. Feng, J. Li, G. Yang, and W. Zheng. VDB-MR: MapReduce-based distributed data integration using virtual database. Future Gener. Comput. Syst. 2010 :26(8) :1418–1425.

N. ABADIE. Formalisation, acquisition et mise en oeuvre des connaissances pour l'intégration virtuelle des bases de données géographiques. Université de Paris-Est, Doctorat Nationale, 2012.

T. Zhao, C. Zhang, M. Wei, and Z.-R. Peng. Ontology-based geospatial data query and integration. Geographic Information Science, Springer, 2008:370–392.

A. Aji, F. Wang, H. Vo, R. Lee, Q. Liu, X. Zhang, and J. Saltz. Hadoop GIS: a high performance spatial data warehousing system over mapreduce. Proceedings of VLDB Endow. 2013:6(11):1009–1020.

N. Berkani, S. Khouri, and L. Bellatreche. Generic methodology for semantic data warehouse design: From schema definition to etl. Intelligent Networking and Collaborative Systems (INCoS), 2012 4th International Conference. 2012:404–411.

T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. Int. J. Hum.-Comput. Stud. 1995:43(5):907–928.

M. Gagnon. Ontology-based integration of data sources. Information Fusion, 2007 10th International Conference. 2007:1-8.

E. Jiménez-Ruiz, B. C. Grau. Logmap: Logic-based and scalable ontology matching. The Semantic Web–ISWC 2011, Springer. 2011:273–288.

S. Mate, F. Köpcke, D. Toddenroth, M. Martin, H. U. Prokosch, T. Bürkle, and T. Ganslandt. Ontology-Based Data Integration between Clinical and Research Systems. PloS One. 2015:10 (1). http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4294641/

A. A. Krisnadhi, K. J. Yingjie Hu, P. Hitzler, R. Arko, S. Carbotte, C. Chandler, M. Cheatham, D. Fils, T. Finin, P. Ji, M. Jones, N. Karima, K. Lehnert, A. Mickle, T. Narock, M. O'Brien, L. Raymond, A. Shepherd, M. Schildhauer, and P. Wiebe. The geolink framework for pattern-based linked data integration. Proceedings of the ISWC 2015 Posters & Demonstrations Track co-located with the 14th International Semantic Web Conference (ISWC-2015), 2015.