



HAL
open science

Applying ontologies to data integration systems for bank credit risk management

Jalil El Elhassouni, Mehdi Bazzi, Abderrahim El Qadi, Mohamed El Haziti

► To cite this version:

Jalil El Elhassouni, Mehdi Bazzi, Abderrahim El Qadi, Mohamed El Haziti. Applying ontologies to data integration systems for bank credit risk management. 2017. hal-01456090v1

HAL Id: hal-01456090

<https://hal.science/hal-01456090v1>

Preprint submitted on 3 Feb 2017 (v1), last revised 26 Nov 2017 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Applying ontologies to data integration systems for bank credit risk management

Jalil ElHassouni^{1*}, Mehdi Bazzi², Abderrahim El Qadi³, Mohamed El Haziti¹

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

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

3TIM Team, High School of Technology - MouayIsmaïl University, Meknes 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 describing credit risk management process and two other locals, the first, describes the credit granting process, and the second presents the concepts necessary for the monitoring of credit system. This paper also presents the technique used for matching between global ontology and local ontologies.

keywords

credit risk management; data integration; ontologies alignment

INTRODUCTION

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

However, the data integration processes across multiple sources is not a simple task because of the semantic heterogeneity encountered in the different levels: metadata, data schemas and data values. This issue is at the heart of many research areas; our work is part of this problem of mapping schema for bank credit risk management. It is to bear in mind the requests of decisions makers and the indicators used in the banking industry to measure the risk of a credit.

The main step in a risk management process is presented 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 allocation credit process to a client and the second presents the concepts necessary for the monitoring of a credit system. Subsequently, we made the alignment of specifics ontologies with the general ontology to facilitate semantic interoperation between applications based on these ontologies.

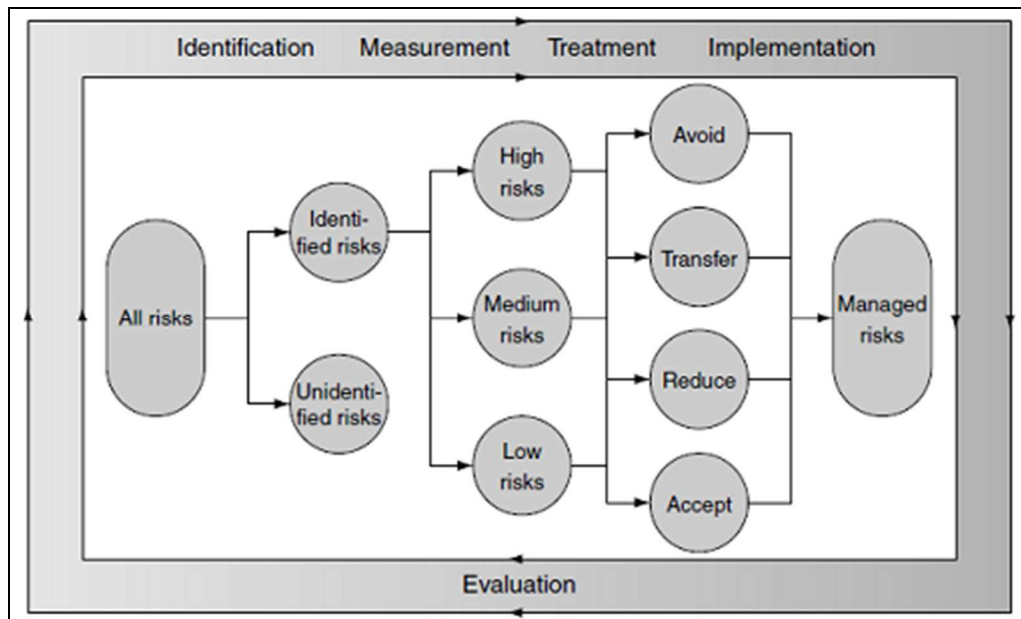


Figure 1. Different steps of a risk management process.

The figure 1 presents different steps of risk management process:

- Identifying and quantifying the risk in order to measure its impact.
- Treat this risk (according to the impact and the strategy of the bank) by transferring, reducing or accepting the risk.
- Once the risk management strategy has been defined, it is 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 existing systems.

The remainder of this paper is organized as follows: in section 1, we provide a literature review. In section 2 we show the process of creating ontologies in the field of bank credit risk. In the last section, we conclude 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: the exploitation of the data is based on the virtual views.
- The Data Warehouse approach (DW): The data is duplicated in the same global schema.

In the mediator approach, the integration of data is done in two steps:

- The integration system generates from a user request as many sub-queries as data sources requested.
- 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, when the user requests are not predictable and when the number of data sources and the number of data 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 request of the user is asked in this data warehouse without access to original data sources. This approach is well suited to store when the information was summarized and annotated or when the user requests can be predefined.

[Calvanese et al, 2001] proposes a method based on a global ontology as intermediate model between DW and sources schemes. [Lamolle and Zerdazi, 2005] provide a mediation model based on XML. [Bakhtouchi et al, 2009] presents a hybrid framework to improve the performance of queries in heterogeneous and autonomous data integration system. [Dean and Ghemawat, 2004] and [Yuan et al, 2010] develop an approach based on MapReduce. Other works use manual or semi-automatic solutions requiring more time and the intervention of domain experts. Furthermore, the two proposals [Aji et al, 2013] and [Berkani et al, 2012] are adapted to support dynamic queries, which can meet the new needs of users. Another important point we notice is the nature of supported data. In the data integration we note that the approaches proposed in [Abadie, 2012], [Zhao et al, 2008] and [Aji et al, 2013] manipulate spatial data while other proposals manipulate alphanumeric data as presented in [Yuan et al, 2010] or ontologies based data as presented in [Berkani et al, 2012].

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]. “Formal” means ontologies should be expressed in a formal and semantic language (RDF, OWL, PLIB...) allowing automatic reasoning. “Explicit” means accepted by all expert community members and “Conceptualization” means an abstract, simplified view of the world that we wish to represent for some purpose.

According to [Gagnon, 2007] the “ontologies integration” characterizes 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.: build a single one that unifies all of them.
- Usage in applications: building an application ontology using one or more ontologies. Existing ontologies are adapted if necessary.

2.1. Allocating Credit to customers (Credit Scoring)

The figure 2 presents the architecture of the system of credit scoring. Indeed, we can divide this process into two phases. The first is about 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 concerns the use of the statistical model to make credit decision. Fixing the cut-off, setting credit limits, pricing and storing the information.

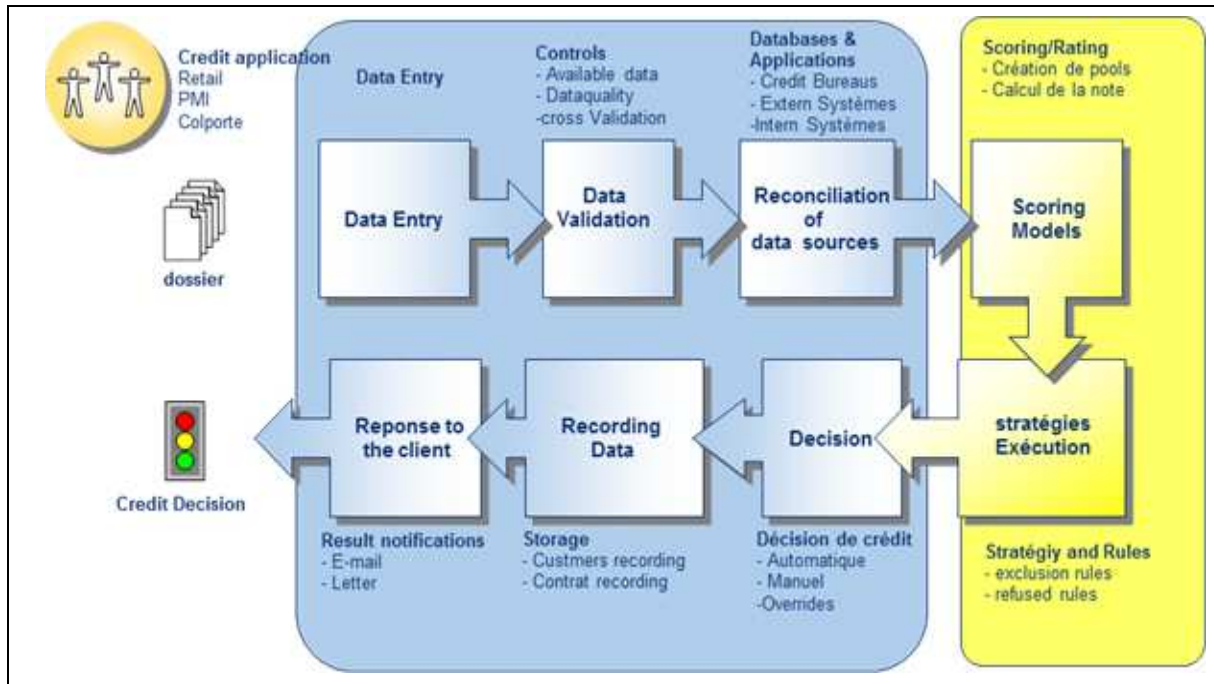


Figure 2. Process of allocating credit to customers (Credit Scoring).

Based on 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 Requires), others are related to the financial profile (Financial Data) and to the personal profile (Socio Demographic) of the borrower while other variables are related to the final decision and to the rating of the borrower (Decision Data). We modeled the credit allocation process on the form of ontology figure 4.

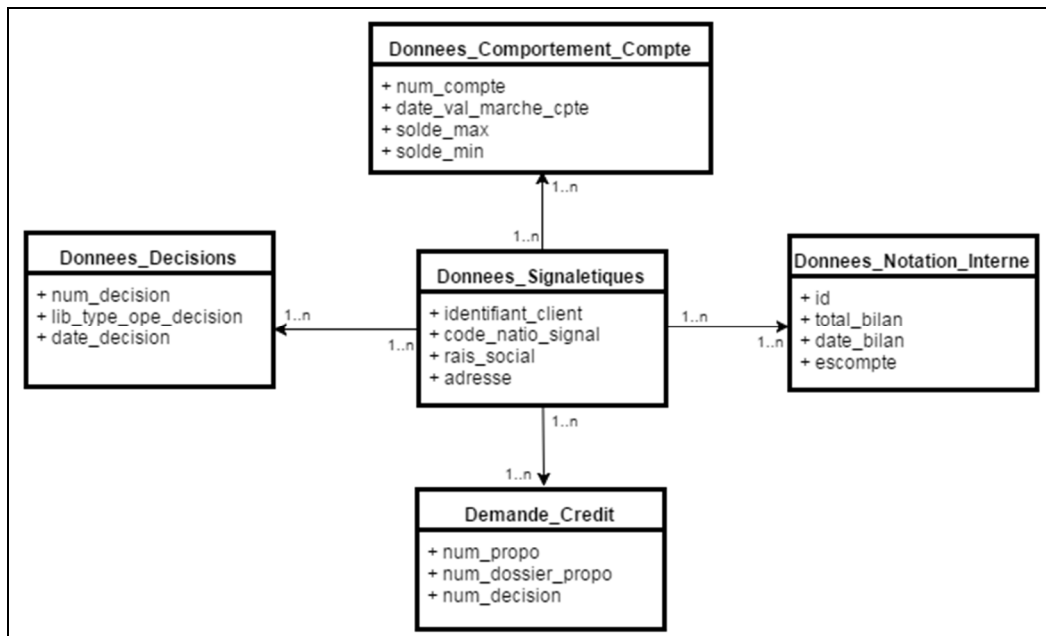


Figure 3. Conceptual model of credit risk management.

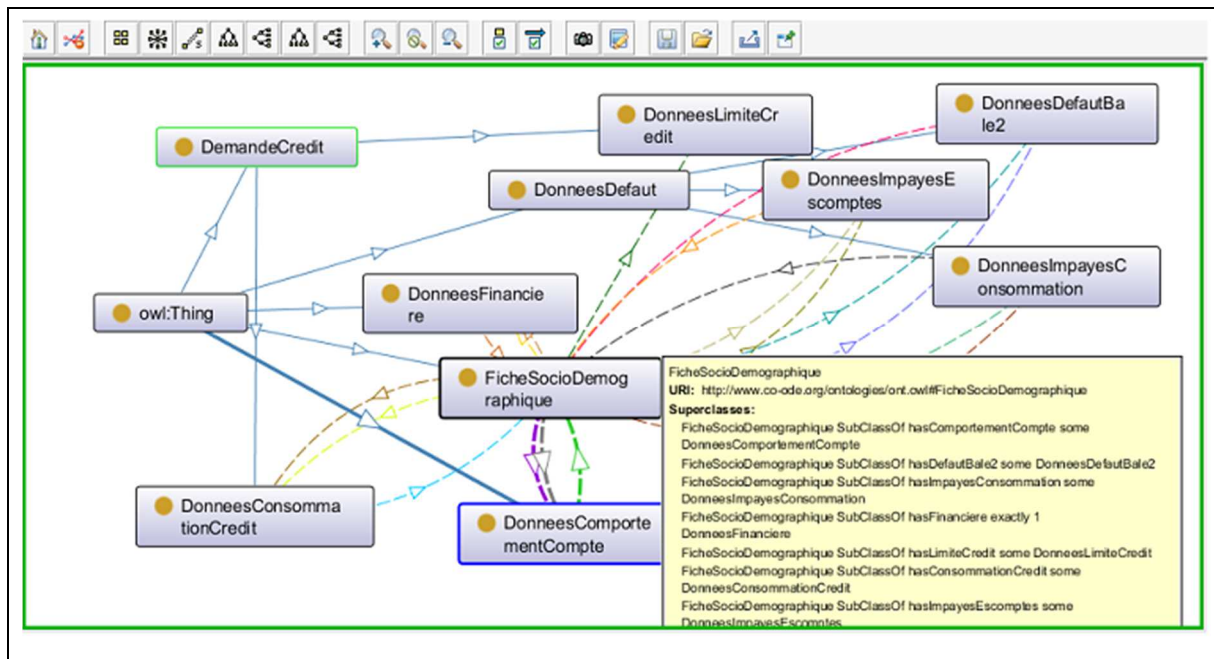


Figure 4. The ontology of the credit allocation process - partial view.

2.2. Ontology for the monitoring of credit

Credit monitoring is related to ‘after credit decision’ collection, analysis of information, and most importantly identification of early warning signs, and ensuring the appropriate decisions are made with regard to changes in the credit risk of a customer.

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 main steps in credit monitoring. We modeled the monitoring of credit process on the form of ontology in figure 6.

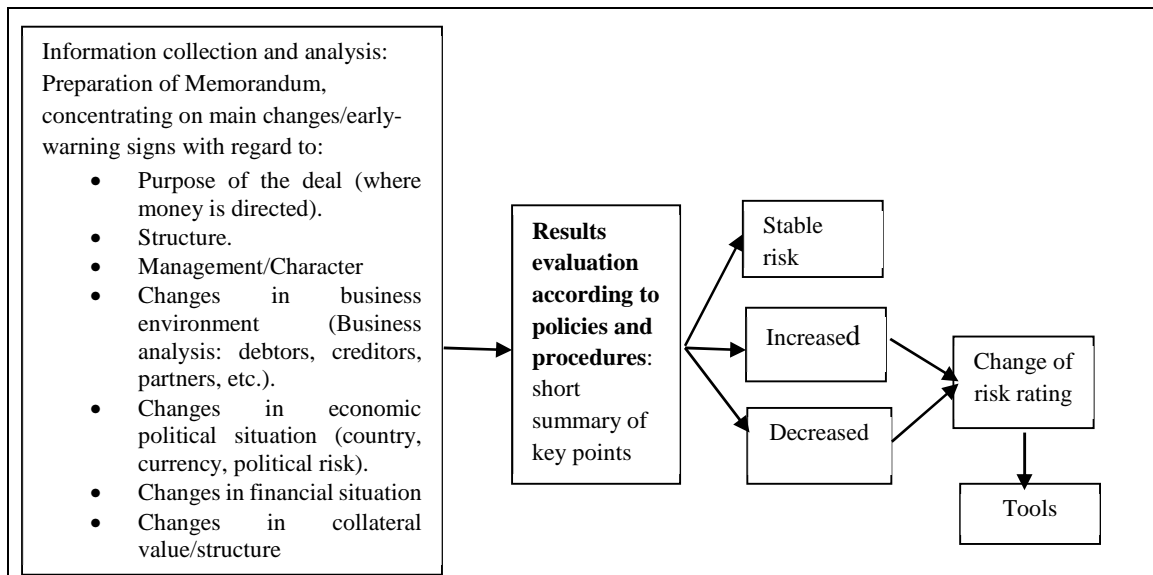


Figure 5. Main steps in credit monitoring.

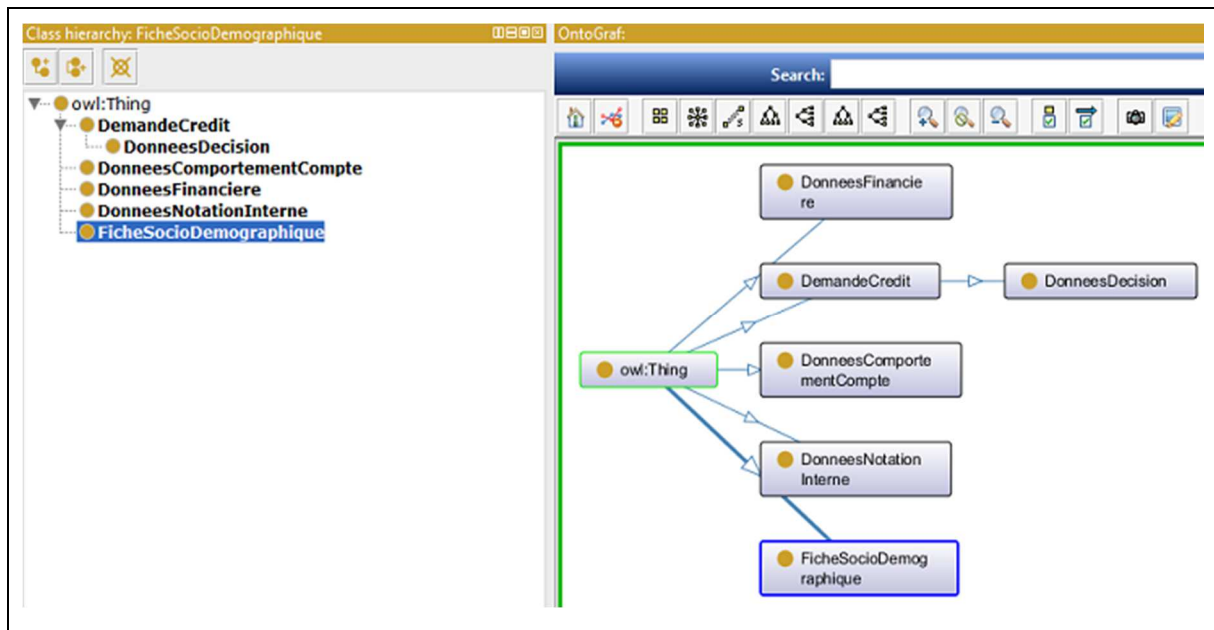


Figure 6. Monitoring of credit process ontology.

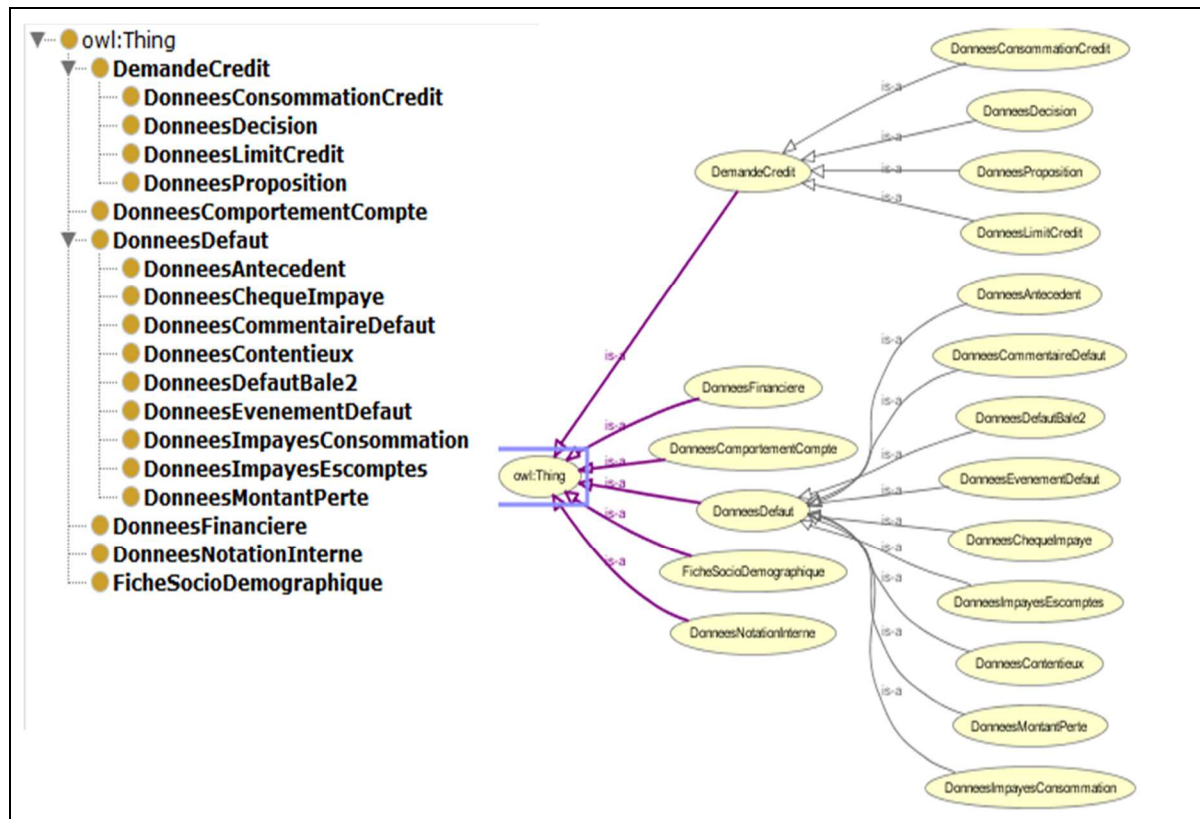


Figure 7. The ontology of the monitoring of credit process - partial view.

2.3. Global process of risk management

Now that we have defined the credit allocation ontology and the monitoring of credit 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 have established several meetings and brainstorming sessions with the banking experts; that is how we came to develop a version described in Figure 7.

```

http://www.co-ode.org/ontologies/ont.owl#DonneesLimitCredit|http://www.co-ode.org/ontologies/ont.owl#DonneesLimiteCredit|=|0.99|CLS
http://www.co-ode.org/ontologies/ont.owl#FicheSocioDemographique|http://www.co-ode.org/ontologies/ont.owl#FicheSocioDemographique|=|0.7|CLS
http://www.co-ode.org/ontologies/ont.owl#DemandeCredit|http://www.co-ode.org/ontologies/ont.owl#DemandeCredit|=|0.9|CLS
http://www.co-ode.org/ontologies/ont.owl#DonneesFinanciere|http://www.co-ode.org/ontologies/ont.owl#DonneesFinanciere|=|0.7|CLS
http://www.co-ode.org/ontologies/ont.owl#DonneesImpayeesConsommation|http://www.co-ode.org/ontologies/ont.owl#DonneesImpayeesConsommation|=|1.0|CLS
http://www.co-ode.org/ontologies/ont.owl#DonneesDefault|http://www.co-ode.org/ontologies/ont.owl#DonneesDefault|=|0.94|CLS
http://www.co-ode.org/ontologies/ont.owl#DonneesComportementCompte|http://www.co-ode.org/ontologies/ont.owl#DonneesComportementCompte|=|0.7|CLS
http://www.co-ode.org/ontologies/ont.owl#DonneesDefaultBale2|http://www.co-ode.org/ontologies/ont.owl#DonneesDefaultBale2|=|1.0|CLS
http://www.co-ode.org/ontologies/ont.owl#DonneesImpayeesEscomptes|http://www.co-ode.org/ontologies/ont.owl#DonneesImpayeesEscomptes|=|1.0|CLS
http://www.co-ode.org/ontologies/ont.owl#DonneesConsommationCredit|http://www.co-ode.org/ontologies/ont.owl#DonneesConsommationCredit|=|1.0|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#hasCodPdtSysImpAmort|http://www.co-ode.org/ontologies/ont.owl#hasCodPdtSysImpAmort|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasRepriseNcBilan|http://www.co-ode.org/ontologies/ont.owl#hasRepriseNcBilan|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasMaterielBilan|http://www.co-ode.org/ontologies/ont.owl#hasMaterielBilan|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasSoldMoyDbAn2|http://www.co-ode.org/ontologies/ont.owl#hasSoldMoyDbAn2|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasSoldMoyDbAn1|http://www.co-ode.org/ontologies/ont.owl#hasSoldMoyDbAn1|=|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#hasAutCharFinBilan|http://www.co-ode.org/ontologies/ont.owl#hasAutCharFinBilan|=|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#hasDtDernEcheImpAmort|http://www.co-ode.org/ontologies/ont.owl#hasDtDernEcheImpAmort|=|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#hasImmoEncBilan|http://www.co-ode.org/ontologies/ont.owl#hasImmoEncBilan|=|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#hasDateCreatDefault|http://www.co-ode.org/ontologies/ont.owl#hasDateCreatDefault|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasRepSubvInvBilan|http://www.co-ode.org/ontologies/ont.owl#hasRepSubvInvBilan|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasNumCpteMrch|http://www.co-ode.org/ontologies/ont.owl#hasNumCpteMrch|=|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#hasVteProdExBilan|http://www.co-ode.org/ontologies/ont.owl#hasVteProdExBilan|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasResultNetBilan|http://www.co-ode.org/ontologies/ont.owl#hasResultNetBilan|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasCodPostSignal|http://www.co-ode.org/ontologies/ont.owl#hasCodPostSignal|=|1.0|DPROP
http://www.co-ode.org/ontologies/ont.owl#hasCodPdtSysImpEsc|http://www.co-ode.org/ontologies/ont.owl#hasCodPdtSysImpEsc|=|1.0|DPROP

```

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

2.4. Ontologies alignment

In the last decade, several studies were released in schema matching area, including alignment ontology. This Alignment Ontology 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 8, using LogMap Ontology Matching Tool [Jiménez-Ruiz and Grau, 2011].


```

//
// Classes
//
////////////////////////////////////

-->
▶<!--.....>
<rdf:Description rdf:about="http://www.co-ode.org/ontologies/ont.owl#DemandeCredit"/>
▶<!--.....>
<rdf:Description rdf:about="http://www.co-ode.org/ontologies/ont.owl#DonneesComportementCompte"/>
▶<!--.....>
<rdf:Description rdf:about="http://www.co-ode.org/ontologies/ont.owl#DonneesConsommationCredit"/>
▶<!--.....>
<rdf:Description rdf:about="http://www.co-ode.org/ontologies/ont.owl#DonneesDefaut"/>
▶<!--.....>
<rdf:Description rdf:about="http://www.co-ode.org/ontologies/ont.owl#DonneesDefautBale2"/>
▶<!--.....>
<rdf:Description rdf:about="http://www.co-ode.org/ontologies/ont.owl#DonneesFinanciere"/>
▶<!--.....>
<rdf:Description rdf:about="http://www.co-ode.org/ontologies/ont.owl#DonneesImpayeesConsommation"/>
▶<!--.....>
<rdf:Description rdf:about="http://www.co-ode.org/ontologies/ont.owl#DonneesImpayeesEscomptes"/>
▶<!--.....>
▼<owl:Class rdf:about="http://www.co-ode.org/ontologies/ont.owl#DonneesLimitCredit">
  <owl:equivalentClass rdf:resource="http://www.co-ode.org/ontologies/ont.owl#DonneesLimiteCredit"/>
</owl:Class>
▼<owl:Axiom>
  <measure rdf:datatype="http://www.w3.org/2001/XMLSchema#double">0.991891891891892</measure>
  <relation rdf:datatype="http://www.w3.org/2001/XMLSchema#string">=</relation>
  ▼<entity1 rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    http://www.co-ode.org/ontologies/ont.owl#DonneesLimitCredit
  </entity1>
  ▼<entity2 rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    http://www.co-ode.org/ontologies/ont.owl#DonneesLimiteCredit
  </entity2>
  <owl:annotatedSource rdf:resource="http://www.co-ode.org/ontologies/ont.owl#DonneesLimitCredit"/>
  <owl:annotatedTarget rdf:resource="http://www.co-ode.org/ontologies/ont.owl#DonneesLimiteCredit"/>
  <owl:annotatedProperty rdf:resource="http://www.w3.org/2002/07/owl#equivalentClass"/>
</owl:Axiom>
▼<!--
  http://www.co-ode.org/ontologies/ont.owl#DonneesLimiteCredit
-->
<owl:Class rdf:about="http://www.co-ode.org/ontologies/ont.owl#DonneesLimiteCredit"/>
▼<owl:Axiom>

```

Figure 9. The alignment file of credit monitoring ontology with the global ontology in text format - partial view.

Figures 8 and 9, show the ‘correct’ mapping founded by LogMap algorithm, we can observe those entities such as classes and data properties in both ontologies that are semantically equivalent and in the similarity between those entities.

Ontologies are created to follow up credits of various customers of the bank, it allows several analyses including:

- The analysis of stability of the elaborated models; such analysis 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 experience in 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, and take into account the requests of decision makers in the process of integration. We also plan to develop a hybrid model for management risk credit bank based on ontologies and MapReduce approach.

References

- Hacid M.S. and Reynaud C. L'intégration de sources de données, dans la revue Information – Interaction – intelligence – (R 13), 2004.
- Calvanese D., De Giacomo G., Lenzerini M., Nardi D. and Rosati R. Data integration in data warehousing, *Int. J. Coop. Inf. Syst.*, vol. 10, no. 03, 2001, 237–271.
- Lamolle M. and Zerdazi A. Intégration de Bases de données hétérogènes par une modélisation conceptuelle XML, *COSI*, vol. 5, 2005, 216–227.
- Bakhtouchi A., L. Bellatreche and Amar Balla, Materializing attributes annotation: A hybrid approach for databases integration, *STIC 2009*.
- Dean J. and Ghemawat S. MapReduce: Simplified Data Processing on Large Clusters, *OSDI 2004*.
- Yuan Y., Wu Y., Feng X., Li J., Yang G. and Zheng W. VDB-MR: MapReduce-based distributed data integration using virtual database, *Future Gener. Comput. Syst.*, vol. 26, no.8, Oct. 2010, 1418–1425.
- Abadie N., Formalisation, acquisition et mise en oeuvre des connaissances pour l'intégration virtuelle des bases de données géographiques, Université de Paris-Est, Doctorat Nationale, 2012.
- Zhao T., Zhang C., Wei M. and Peng Z.R. Ontology-based geospatial data query and integration, in *Geographic Information Science*, Springer, 2008, 370–392.
- Aji A., Wang F., Vo H., Lee R., Liu Q., Zhang X. and Saltz J. Hadoop GIS: a high performance spatial data warehousing system over mapreduce, *Proc. VLDB Endow.*, vol. 6, no. 11, 2013, 1009–1020.
- Berkani N., Khouri S. and Bellatreche L. Generic methodology for semantic data warehouse design: From schema definition to etl, in *Intelligent Networking and Collaborative Systems (INCoS)*, 2012 4th International Conference on, 2012, 404–411.
- Gruber T. R. Toward principles for the design of ontologies used for knowledge sharing, *Int. J. Hum.-Comput. Stud.* vol. 43, no 5, 1995, 907–928.
- Gagnon M. Ontology-based integration of data sources, in *Information Fusion*, 2007 10th International Conference on, 2007, 1-8.
- Jiménez-Ruiz E. and Grau B. C. Logmap: Logic-based and scalable ontology matching, in *The Semantic Web–ISWC 2011*, Springer, 2011, 273–288.