ROYAUME DU MAROC
Université Mohammed V - RABAT - جامعة محمد الخامس – الرباط –
Faculté des sciences
كلية العلوم

CENTRE D'ETUDES DOCTORALES - SCIENCES ET TECHNOLOGIES

N° d'ordre : 3426

THESE

En vue de l'obtention du : ***DOCTORAT***

Structure de Recherche : Laboratoire de Recherche en Informatique et Télécommunications
Discipline : Sciences de l'ingénieur
Spécialité : Informatique

Présentée et soutenue le : 06/02/2021 par :

Jalil ELHASSOUNI

Towards ontology-enhanced knowledge-based systems using ontology design patterns: application in credit risk management

JURY

Mme. Salma MOULINE	PES, Université Mohammed V - Rabat, Faculté des Sciences	Présidente
M. ElHoussaine ZIYATI	PH, Université Hassan II - Casablanca, Ecole Supérieure de Technologie	Rapporteur/Examinateur
M. Mohamed LAZAAR	PH, Université Mohammed V - Rabat, Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes	Rapporteur/Examinateur
M. Rachid ALAOUI	PH, Université Mohammed V - Rabat, Ecole Supérieure de Technologie de Salé	Rapporteur
Mme. Khadija RHOULAMI	PH, Université Mohammed V - Rabat, Faculté des Lettres et des Sciences Humaines	Examinatrice
M. Abderrahim EL QADI	PES, Université Mohammed V - Rabat, École Nationale Supérieure d'Arts et Métiers	Co-Directeur de Thèse
M. Mohammed EL HAZITI	PES, Université Mohammed V - Rabat, Ecole Supérieure de Technologie de Salé	Directeur de Thèse

Année Universitaire : 2020-2021

✉ Faculté des Sciences, avenue Ibn Battouta, BP. 1014 RP, Rabat –Maroc
☎ 00212(0)5 37 77 18 76 🖷 00212(0)5 37 77 42 61 ; http://www.fsr.um5.ac.ma

ROYAUME DU MAROC
Université Mohammed V - RABAT - جامعة محمد الخامس - الرباط -
Faculté des sciences
كلية العلوم

CENTRE D'ETUDES DOCTORALES - SCIENCES ET TECHNOLOGIES

N° d'ordre : 3426

THESE

En vue de l'obtention du : ***DOCTORAT***

Structure de Recherche : Laboratoire de Recherche en Informatique et Télécommunications
Discipline : Sciences de l'ingénieur
Spécialité : Informatique

Présentée et soutenue le : 06/02/2021 par :

Jalil ELHASSOUNI

Towards ontology-enhanced knowledge-based systems using ontology design patterns: application in credit risk management

JURY

Mme. Salma MOULINE	PES, Université Mohammed V - Rabat, Faculté des Sciences	Présidente
M. ElHoussaine ZIYATI	PH, Université Hassan II - Casablanca, Ecole Supérieure de Technologie	Rapporteur/Examinateur
M. Mohamed LAZAAR	PH, Université Mohammed V - Rabat, Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes	Rapporteur/Examinateur
M. Rachid ALAOUI	PH, Université Mohammed V - Rabat, Ecole Supérieure de Technologie de Salé	Rapporteur
Mme. Khadija RHOULAMI	PH, Université Mohammed V - Rabat, Faculté des Lettres et des Sciences Humaines	Examinatrice
M. Abderrahim EL QADI	PES, Université Mohammed V - Rabat, École Nationale Supérieure d'Arts et Métiers	Co-Directeur de Thèse
M. Mohammed EL HAZITI	PES, Université Mohammed V - Rabat, Ecole Supérieure de Technologie de Salé	Directeur de Thèse

Année Universitaire : 2020-2021

Dedication

To my parents, you have always been an example. Your outlook on life, your thoroughness and your affection made of me the man that I am. I can never be grateful enough for you. My respect and my love for you will remain forever.

To my brothers, Amine and Idriss.

To my nephew Adam.

Acknowledgements

"One can pay back the loan of gold, but one dies forever in debt to those who are kind." -- Malayan proverb

This dissertation titled "Towards ontology-enhanced knowledge-based systems using ontology design patterns: application in credit risk management" was prepared at the Research Laboratory in Computer Science and Telecommunications, associated unit with CNRST (URAC No 29), Faculty of Sciences, Mohammed V University in Rabat under the supervision of both Pr. Mohammed El Haziti and Pr. Abderrahim El Qadi.

This work bears the name of one individual, but in fact it is the result of the good efforts of many. It is with much pleasure and gratitude that I am thankful first and foremost to my dissertation advisor Mr. Mohammed El Haziti, professor at the Sale Higher School of Technology, and co-advisor Mr. Abderrahim El Qadi, professor at the Higher National School of Arts and Crafts in Rabat, who have been intellectually challenging and inspiring as well as extremely patient, supportive and humble.

I would like to express my sincere gratitude to the president of the jury Ms. Salma Mouline, professor at Rabat Faculty of Sciences for contributing to the examination and evaluation of this dissertation as an examinator from Mohammed V University of Rabat.

I would like to equally extend my thankfulness to Mr. ElHoussaine Ziyati, professor at the Casablanca Higher School of Technology for his contribution to the examination and evaluation of this dissertation as an examinator and reporter from Hassan II University of Casablanca.

I am also grateful to Mr. Mohamed Lazaar, professor at the Rabat Higher National School of Computer Science and Systems Analysis, who contributed to the examination and evaluation of this dissertation as an examinator and reporter from Mohammed V University of Rabat.

My thanks are also due to Mr. Rachid Alaoui, professor at the Sale Higher School of Technology, for his contribution as a reporter from Mohammed V University of Rabat.

I would also like to thank Ms. Khadija Rhoulami, professor at Rabat Faculty of

Letters and Humanities, for contributing to the examination and evaluation of this dissertation as an examinator from Mohammed V University of Rabat.

I would like to thank Mr. Mehdi Bazzi for his collaboration and contribution to the realization of this work. Many thanks are also due to Dr. Yasser El Madani El Alami for his diligence and invaluable support and feedback.

I would also like to thank my friends Mohamed Saoudi, Othman Ouaarab, and Abdellah Ajeraam who have encouraged me all the way through.

This acknowledgment would not be complete without proper thanks to the Faculty of Sciences in Rabat, the Department of Physics, and the LRIT Laboratory for sponsoring this dissertation. My thanks also to the administrative staff who have made my experience at this faculty memorable and beneficial.

Equally appreciated is the support I received from the Urban Agency of Casablanca "Agence Urbaine de Casablanca" for realizing one of my dreams through pursuing my doctoral studies.

Last but not least, I am also grateful to the "Association socio-culturelle du Bassin Méditerranéen" and to all the people who have dedicated themselves to managing and providing a learning space for us.

Author declares that there is no conflict of interest.

Abstract

Recently, the amount of data generated on a daily basis has exponentially exploded. Tremendous efforts have been undertaken to provide the industry and research fields with expanding technologies for facilitating the collection, treatment, and sharing of massive data. This poses unprecedented challenges for the use of these technologies, especially in the credit risk management. The greatest challenge lies in collecting, retrieving, reusing and sharing this knowledge on one hand, and in creating new values for improving the competitiveness of financial institutions on the other hand. In the aftermath of the 2008 financial crisis, banking systems have shown a precarious fragility particularly in data architecture and information technology infrastructure. Moreover, inadequacy or non–existence of common vocabulary has created a gap in common semantics. Several solutions have been proposed, but the most efficient approach to respond to these issues is the use of ontologies. Ontologies are a natural fit for data knowledge representation, and data storing, retrieving, integrating, and reasoning. This work used the ontology for modeling the credit risk scorecard. We opted for the Krisnadhi & Hitzler methodology based on Ontology Design Patterns, which is a worked and tested example, to develop credit risk scorecard and applicant ontologies. The methodology supports collaborative construction and allows for reusability. We linked the credit risk scorecard to credit risk scorecard objectives and credit risk scoring decision support tool. Our credit risk scorecard ontology is then used to identify the corresponding tasks, subtasks, roles, actors, and resources of the affected business. Our ontology is validated using the methods and tools already tested. Firstly, we make the diagnosis and repair of the ontology for quality validation. Secondly, we reason the logical consistency of the ontology. Thirdly, we make queries answering for usability validation of the ontology. Our ontology is a shareable ontology that can be understood both by humans and computers alike. The major aim of this dissertation is to develop a rich, expandable, and re-usable credit risk scorecard based on ontology design patterns and in compliance with BCBS 239. The principal result of this work is laying the corner stone for the credit risk management platform-based ontology, which remains open to the extent that it allows for constant accumulation of tacit knowledge and efficient polymerization of explicit knowledge.

Keywords: Ontology, Ontology Design Patterns, Knowledge-based systems, Credit risk scorecard, Credit risk management.

Résumé

La quantité de données générées quotidiennement explose, de nos jours, à un rythme exponentiel. Des efforts immenses ont été déployés pour fournir à l'industrie et aux domaines de recherche, des technologies en expansion afin de faciliter la collecte, le traitement et le partage de données massives. Ce qui engendre des défis sans précédent quant à l'utilisation de ces technologies, particulièrement dans la gestion du risque de crédit. Le plus grand défi réside dans la collecte, la récupération, la réutilisation et le partage de ces connaissances d'une part ainsi que dans la création de nouvelles connaissances pour améliorer la compétitivité des institutions financières d'autre part. Au lendemain de la crise financière de 2008, les systèmes bancaires ont fait preuve d’une grande faiblesse notamment dans l'architecture des données et les infrastructures informatiques. En plus, l'insuffisance ou la non-existence d'un vocabulaire commun a créé une lacune dans la sémantique commune. Plusieurs solutions ont été proposées pour résoudre ces problèmes, mais l'approche basée sur l’utilisation des ontologies reste la plus efficace. Les ontologies sont une solution naturelle pour la représentation de la connaissance, le stockage, la recherche, l'intégration et l’analyse des données « reasoning data ». Ce travail a utilisé l'ontologie pour modéliser le modèle de notation « scorecard » du risque de crédit. Nous avons opté pour la méthodologie de Krisnadhi & Hitzler basée sur des patrons de conception d’ontologie « Ontology Design Patterns », une méthodologie qui a prouvé son efficacité pour le développement et le test des cas réels, afin de développer les ontologies du modèle de notation du risque de crédit et du demandeur. La méthodologie soutient la construction collaborative et permet la réutilisation. Nous avons fait la liaison du modèle de notation du risque de crédit avec ces objectifs et l'outil d'aide à la décision du risque de crédit. Notre ontologie de modèle de notation du risque de crédit est ensuite utilisée pour identifier les tâches, sous-tâches, rôles, acteurs et ressources correspondants aux domaines connexes. Notre ontologie a été validée à l'aide des méthodes et outils déjà testés. Premièrement, nous avons fait le diagnostic et la réparation pour la validation de la qualité de l'ontologie. Deuxièmement, nous avons travaillé sur le raisonnement et la cohérence logique de l'ontologie. Troisièmement, nous avons formulé des requêtes pour valider l'utilisation de l'ontologie. Notre ontologie est une ontologie partageable, compréhensive à la fois par les humains et les ordinateurs. L'objectif principal de cette thèse est de développer un modèle de notation du risque de crédit riche, extensible et réutilisable basée sur des patrons de conception d'ontologie et en conformité avec le BCBS 239. Le résultat principal de ce travail est de poser le premier bloc de la plateforme de gestion du risque de crédit basée sur l’ontologie, qui reste ouverte dans la mesure où elle permet une accumulation constante de connaissances tacites et une polymérisation efficace des connaissances explicites.

Mots-clés : Ontologie, Patrons de conception d'ontologie, Systèmes à base de connaissances, Modèle de notation du risque de crédit, Gestion du risque de crédit.

ملخص

في الآونة الأخيرة، ازداد حجم البيانات التي يتم تداولها يوميا بشكل كبير. لقد تم بذل جهود جبارة لتزويد الصناعة ومجالات البحث بتقنيات في تطور مُطَّرد لتسهيل جمع ومعالجة ومشاركة البيانات الضخمة، مما يفرض تحديات غير مسبوقة على استخدام هذه التقنيات خاصة في إدارة مخاطر الائتمان. ويكمن التحدي الأكبر في جمع هذه المعرفة واسترجاعها وإعادة استخدامها ومشاركتها من جهة، وفي خلق قيم جديدة لتحسين القدرة التنافسية للمؤسسات المالية من جهة أخرى. لقد أظهرت الأنظمة البنكية هشاشة محفوفة بالمخاطر خاصة في بنية البيانات والبنية التحتية لتكنولوجيا المعلومات عقب الأزمة المالية لعام 2008. علاوة على ذلك، أدى عدم ملائمة أو عدم وجود معجم موحد إلى خلق فجوة في الدلالات الموحدة. وقد تم اقتراح واستخدام عدة حلول، لكن يبقى الحل الأكثر فعالية لمعالجة هذه القضايا هو استخدام الأنطولوجيا. تعتبر الأنطولوجيا الحل المناسب لتمثيل البيانات وتخزينها، واسترجاعها، ودمجها، واستدلالها. استخدم هذا العمل الأنطولوجيا لنمذجة نظام تنقيط «Scorecard» المخاطر الائتمانية. اخترنا منهجية «Krisnadhi & Hitzler» استنادًا إلى «Ontology Design Patterns» وهي منهجية عملية ومختبرة لتطوير واختبار الحالات الحقيقية، لتطوير نظام تنقيط مخاطر القرض البنكي وأنطولوجيا لمقدم الطلب. تدعم هذه المنهجية التشييد التعاوني وتسمح بإعادة الاستخدام. لقد ربطنا نظام تنقيط مخاطر الائتمان وأهدافها بأداة دعم قرارات مخاطر الائتمان. كما استخدمنا الأنطولوجيا الخاصة بنظام تنقيط مخاطر الائتمانية لتحديد المهام والمهام الفرعية والأدوار والجهات الفاعلة وموارد المجالات ذات الصلة، وتحققنا من فعالية الأنطولوجيا الخاصة بنا باستخدام الأساليب والأدوات التي تم اختبارها من قبل. أولاً، قمنا بتشخيص وإصلاح للتحقق من جودة الأنطولوجيا. ثانيًا، عملنا على الاستدلال في التناسق المنطقي للأنطولوجيا. ثالثًا، قمنا بصياغة مجموعة من الاستفسارات للتحقق من قابلية استخدام الأنطولوجيا. هذه الأنطولوجيا قابلة للمشاركة، كما يمكن أن يفهمها كل من البشر وأجهزة الكمبيوتر على حد سواء. الهدف الرئيسي من هذه الأطروحة هو تطوير نظام تنقيط مخاطر الائتمانية غني وقابل للتوسيع وإعادة الاستخدام بناءً على أنماط تصميم الأنطولوجيا وبتوافق مع مبادئ « BCBS 239». تتجلى الثمرة الرئيسية لهذا العمل هي وضع اللبنة الأولى لمنصة إدارة مخاطر الائتمانية قائمة على الأنطولوجيا، بحيث تظل مفتوحة إلى الحد الذي تسمح فيه بالتراكم المستمر للمعرفة الضمنية والبلمرة الفعالة للمعرفة الصريحة.

الكلمات المفتاحية: الأنطولوجيا، أنماط تصميم أنطولوجيا، أنظمة قائمة على المعرفة، نموذج تنقيط مخاطر الائتمانية، إدارة مخاطر الائتمان البنكي.

Résumé détaillé

Récemment, la quantité de données générées quotidiennement par les nouvelles technologies telles que les réseaux sociaux et l'Internet des Objets (IoT) explose à un rythme exponentiel. Les ontologies promettent d'utiliser au mieux ces informations (stockage, récupération, intégration et analyse des données). L'utilisation d'ontologies permettra aux données d'être accessibles, partageables et réutilisables. Ce travail tente de mettre en place ces objectifs et de les appliquer dans le secteur bancaire, notamment dans la gestion du risque de crédit.

L'objectif principal de cette thèse réside dans le développement du modèle de notation « scorecard » du risque de crédit riche, extensible et réutilisable basé sur les patrons de conception d'ontologie tout en répondant aux défis spécifiques auxquels la gestion du risque de crédit a été confrontée après la crise financière de 2008, ainsi que le coût1 très chère de la mise en œuvre du CBCB 239.

Nous avons commencé par le recensement les plus importants rôles des ontologies dans le développement des systèmes à base de connaissances fin de justifier le choix de leurs utilisations. Deuxièmement, nous avons fait une étude comparative des méthodes de développement d'ontologies afin de choisir la méthode la plus appropriée. Troisièmement, pour développer les ontologies du modèle de notation du risque de crédit et les candidats de façon correctes, fiables et efficaces on a fait appel à notre expertise dans les bases de données et l'ingénierie des ontologies et une équipe d'experts dans le domaine de crédit bancaire. La mise en œuvre desdites ontologies s'est basée sur un processus de modélisation basé sur ODPs. Quatrièmement, afin de développer et de mettre en œuvre nos ontologies, nous avons suivi les étapes ci-dessous :

1) Collection et extraction des données: ‘German credit data'[2].
2) Extraction des données du scorecard, ce dernier est le résultat de la modélisation du score de crédit par régression logistique appliquée aux ‘German credit data'.
3) Synthétisation et analyser des données pour développer un modèle de notation du risque de crédit riche, extensible et réutilisable à l'aide des ODPs et en

---

[1]Selon le rapport McKinsey de 2015, on estime que la mise en œuvre des principes du BCBS 239 coûterait environ 230 millions USD par G-SIB et 75 millions USD par D-SIB.

[2]https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data)

conformité avec BCBS 239.

La contribution principale de cette thèse est de développer une solution qui permettra de faire une meilleure théorie du contenu et une représentation du domaine du modèle de notation du risque de crédit. Elle aidera les institutions financières à développer une plateforme conforme aux principes BCBS 239. Ainsi, elle permettra aux solutions destinées aux utilisateurs développées à base des mécanismes d'IA (par exemple, machine learning et deep learning) de mieux évaluer les potentiels scénarios de cas futurs, prédire le comportement de manière sans précédent et extraire les informations significatives de données. Cette solution créera un vocabulaire commun qui facilitera la mise en œuvre, la réutilisation de la plate-forme et le partage d'informations, à la fois inter et intra les institutions financières.

Nous avons utilisé la méthodologie Krisnadhi & Hitzler basée sur les patrons de conception d'ontologie (ODPs) pour modéliser le modèle de notation du risque de crédit afin d'améliorer la gestion du risque de crédit. La modélisation de notre ontologie reliera directement le modèle de notation du risque de crédit à ces objectifs et à l'outil d'aide à la décision du risque de crédit. Notre proposition est extensible et vise à satisfaire divers besoins. Elle fournit des éléments modulaires, réutilisables et remplaçables, et simplifiera la publication des données et rendra les structures graphiques intuitives, facilitant ainsi leurs extensibilité et réutilisation. Le principal résultat de ce travail est de mettre les bases de la plateforme de gestion du risque de crédit basée sur l'ontologie.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AI : *Artificial Intelligence*
AMOD : *Agile Methodology for Ontology Development*
AS : *Application Scoring*
BCBS : *Basel Committee on Banking Supervision*
BFO : *Basic Formal Ontology*
BS : *Behavioral Scoring*
CGs : *Conceptual Graphs*
CP : *Content OP*
CQs : *Competency Questions*
DLs : *Description Logics*
DOLCE : *Descriptive Ontology for Linguistic and Cognitive Engineering*
DSSs : *Decision Support Systems*
EG : *existential graphs of Charles Sanders Peirce*
FIBO : *Financial Industry Business Ontology*
GO : *Gene Ontology*
IO : *Internet of Things*
IR : *Information Retrieval*
IRS : *Information Retrieval Systems*
IT : *Information Technology*
KBSs : *Knowledge Base Systems*
KI : *Knowledge Integration*
KR&R : *Knowledge Representation and Reasoning*
KS&R : *Knowledge Sharing and Reusing*
ML : *Machine Learning*
NN : *Neural Network*
OBDA : *Ontology-Based Data Access*
OBDI : *Ontology-Based Data Integration*
OOPS : *OntOlogy Pitfall Scanner*
OWA : *Open-world assumption*
OWL : *Ontology Web Language*
RDF : *Resource Description Framework*
SNs : *Semantic Networks*
SUMO : *Suggested Upper Merged Ontology*
TOVE : *Toronto Virtual Enterprise*
UIs : *User Interfaces*
URI : *Uniform Resource Identifier*
W3C : *World Wide Web Consortium*

Chapter 1: Introduction

"Data is the new oil"

1. 1.Problematic and Research Statement

In the last decade, the amount of data generated on a daily basis has exponentially exploded. Numerous efforts have been undertaken to provide the industry and research fields with expanding technologies for facilitating the collection, treatment, and sharing of massive data. This poses unprecedented challenges for the use of these technologies, especially in the banking sector. Therefore, the central bank governors of the G10 and Switzerland created the Basel Committee on Banking Supervision (BCBS)[3] in order to assist them in monitoring and exchanging data tasks based on common approaches and standards. BCBS reported that one of the reasons for the financial crisis of 2008 was associated with problems in data architecture and information technology infrastructure[1]. Moreover, inadequacy or non–existence of common vocabulary has created a gap in common semantics. As a consequence, for many banking systems, the assertions or decisions that depend on the data cannot be reliable.

During the last few decades, credit risk scorecards (see Table 4.1) have been widely used in the banking area due to the proliferation of credit applications for different bank products, including applicants' loan [2]. The credit scoring is a statistical tool which allows banks to determine the risk level of an applicant or a customer. Credit risk management has become more important and complex over the years especially after the 2008 financial crisis. To ensure better management, financial institutions have to achieve some goals: (1) implement regulatory norms and more stringent compliance and reporting requirements, and (2) reduce the cost and the time of data management.

In order to enhance risk management and decision-making processes at financial institutions, the BCBS outlined a set of 14 principles (BCBS 239) [1].

In the 2010 talk, Gov 2.0 Expo in Washington, D.C., Tim Berners-Lee suggested awarding stars for any kind of sharing data [3]. Therefore, governments earn three stars for sharing data on the web a nonproprietary format, for stars on putting it in linked data and a full five stars for connecting data to other data. Following the recommendations of BCBS and Tim Berners-Lee, an ontology-based approach has been proposed to assist different financial institutions in standardizing, monitoring, and exchanging data in the credit risk field.

Recently, multiple companies have invested in many areas of Artificial Intelligence (AI);

[3]https://www.bis.org/bcbs/

e.g., machine learning, deep learning, big data and the Internet of Things (IO) to enhance their performance and promote the evolution of the intelligence community. However, "there is no machine intelligence without (knowledge) representation." [4] In other words, as good as the mechanism is, it cannot do without a good content theory of the domain it deals with. Ontology is the most suitable solution to do that as it provides knowledge about the objective realities in the domain of interest and also how to achieve various objectives.

The best way to respond to these issues rests with the development of ontologies. Ontologies are a natural fit for data storing, retrieving, integrating, and reasoning. The use of ontologies to solve the above problems is very promising in this regard.

In recent years, ontologies have been increasingly used in industry and research fields for different purposes. The development of semantic domain ontology can help in reducing common problems and ambiguities linked to tag based systems [5]. Ontologies representing the domain knowledge have been used to guide the design of the application and to supply the system with semantic technologies [6]–[9]. Ontologies are beneficial when they are used in decision support systems [10].

The modeling of credit risk scorecard ontology will directly link credit risk scorecard to credit risk scorecard objectives and credit risk scoring decision support tool. The ontology is then used to identify the corresponding tasks, subtasks, roles, actors, and resources of the affected business.

The development of ontology is an important step in the development of knowledge base systems. The *Gruninger and Fox* methodology follows stage based model. A stage based model approach is best used when the requirements and scope of ontology are well identified. The methodology supports reusability, allowing ontology engineers to make use of existing knowledge resources and reducing the ontology development efforts and time. For the degree of application dependency, the methodology opts for a semi-dependent approach. However, the methodology does not offer any support for collaborative construction and interoperability. Therefore, using this methodology to develop ontology will restrict the collaborative construction and geographical location, plus it will complicate the communication and sharing knowledge in neighboring domains. Also, the methodology does not provide life cycle recommendations. As for identifying concepts, the methodology uses the middle-out strategy, but it provides only some details for its employed techniques and activities. Unlike the *Gruninger and Fox*, the *METHONTOLOGY* methodology provides sufficient details regarding the deployed techniques, but it does not allow for collaborative construction and interoperability. Similar to these methods, the Basic Formal Ontology (*BFO)* method does not allow for collaborative construction as well, though it supports interoperability. The problem; however, with the *BFO* is that it adopts the top-down approach. While the top-down approach is the ideal approach for ontology development, it runs the risk of remaining highly abstract and difficult to be adapted to and reused in specific domains. In addition to its lack of support for collaborative construction and interoperability, the Agile Methodology for Ontology Development *(AMOD)* method also lacks the reusability support. The *AMOD* methodology resembles in many ways *Gruninger and Fox* except for life cycle recommendation and the degree of application dependency. Kendall and McGuinness methodology stands out from the other methodologies by adopting the developer's consent strategy for identifying concepts. Regardless of being tested, on the whole, it remains vague as to how to define the terminology, collaborative construction, and the tools deployed in ontology development process. The *NeOn* methodology proposes nine flexible ontology-building scenarios based on the reuse and reengineering of ontological, non-ontological, and ODPs resources. Theoretically, it exceeds greatly the methodology employed in this work, but the main weaknesses of this methodology are: (i) it does not explicitly and sufficiently detail the extraction of terminologies, (ii) and its application time consuming [11]. Krisnadhi & Hitzler methodology shares with the previous two methodologies (*Kendall & McGuinness and NeOn*) the fact that they are all based on worked and tested examples; however, this methodology is very practical and much more detailed in a step-by-step manner- which is lacking in the NeOn methodology- for those who wish to develop an ontology based on ODPs. In fact, the main strength of this methodology is that is relies on ODPs for modeling ontologies, which permits the ontology to be reused and expanded, and prevents it from being too specific.

The implementation of credit risk scorecard and applicant ontologies based on ODPs as a stepping stone towards a more complex and sophisticated framework for credit risk management and decision making processes in financial institutions. The credit risk scorecard and applicant ontologies are extended, enriched using logical axioms, validated, distributed in Ontology Web Language (OWL) files, and checked in the test cases using SPARQL.

Semantic technology, "semantic computing", refers to the Semantic Web and its related technologies including RDF, RDFS, OWL, SPARQL and graph database. It focuses on deriving knowledge from data. It captures structures and leverages data based on its context and meaning within a specific knowledge domain. When combined with ontology, semantic technology will enable financial institutions to achieve their goals in credit risk management cited above.

1. 2.Research Objectives, Methodology, and Contributions

1.2.1.Research Objectives

Recently, there has been a vast amount of information generated by the new technologies such social media and Internet of Things (IoT). Ontologies hold the promise to make the best use of these loads of information (data storing, retrieving, integrating, and reasoning). The use of ontologies will permit data to be accessible, shared, and re-used. This research attempts to put these objectives in place and apply them in the banking sector, namely in credit risk management.

The major aim of this dissertation lies in the development of a rich, expandable, and re-usable credit risk scorecard based on ontology design patterns and in compliance with BCBS 239. One of the motivations of this research is to respond to the specific challenges which credit risk management has faced after the 2008 financial crisis as well as the extremely high cost[4] of BCBS 239 implementation.

1.2.2.Research Methodology

The methodology of this research is outlined as follows:

First, we conducted a review of the most prominent roles of ontologies in information knowledge systems to justify the choice of using ontologies. Second, we compared a list of the most used methods for ontology development in order to choose the most appropriate method for ontology engineering, Krisnadhi & Hitzler methodology. Third, to develop the credit risk scorecard and applicant ontologies, we gathered a team of experts: domain expert, participants specializing in database, and ontology engineer working on ODP-based modeling process. Fourth, in order to develop and implement our ontologies, we followed the steps below:

[4]According to the 2015 McKinsey report, it is estimated that the implementation of BCBS 239 principles would cost approximately 230 million USD per G-SIB and 75 million USD in each D-SIB.

1) Collecting and extracting data sets from the website: 'German credit data'[5].
2) Extracting data sets from the scorecard which is the result of credit scoring modeling by logistic regression applied to 'German credit data'.
3) Synthesizing and analyzing the data from both sets to develop a rich, expandable, and re-usable credit risk scorecard using ODPs and in compliance with BCBS 239.

1.2.3.Research Contributions

The main contribution of this dissertation is to develop a solution which will permit to make a good content theory and domain representation of the credit risk scorecard. However, having an appropriate representation of the knowledge is only part of the development of our system. Therefore, we use the ontology-based credit risk scorecard and ontology-based applicant to provide the proper knowledge for the right people in order to help them make the optimal decisions. First, it will help financial institutions to develop a platform which conforms to the principles of BCBS 239. Second, it will enable user-ends solutions made available by AI mechanisms (e.g., machine learning and deep learning) to better asses potential future case scenarios, predict behavior in unprecedented ways, and extract meaningful information from the data. Third, it will create a common vocabulary which will facilitate the implementation and re-usability of platform and sharing of information, at both inter- and intra-levels in financial institutions.

We used the Krisnadhi & Hitzler methodology based on ODPs for modeling the credit risk scorecard to improve credit risk management. The modeling of our ontology will directly link credit risk scorecard to credit risk scorecard objectives and credit risk scoring decision support tool. Our proposal is expandable and aims at satisfying various needs. It provides modular, reusable, replaceable pieces, and it will make the data publication simpler and graph structures intuitive, thus, making its reusability and expandability easier. The principal result of this work is laying the cornerstone for the credit risk management platform-based ontology.

1. 3.Thesis Layout and Brief Overview of Chapters

This thesis is composed of five chapters and conclusion and future work, briefly described as follows:

Chapter 2: This chapter deals with the concept of ontology and its roles in solving the

[5]https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data)

problems of data storing, retrieving, integrating, and reasoning. It reviews the relevant literature on the topic and highlights the potential of ontologies in solving the persistent problems resulting from using huge amounts of information generated by the new technologies.

Chapter 3: This chapter illustrates the ontology engineering process. Although the use of ontologies in both academic and industrial fields is quite popular, many ontology projects have failed primarily due to the poorly specified, underspecified, or lack of requirements and evaluation criteria. To address this issue, we propose an ontology development method based on ODPs, which supports collaborative construction and allows for reusability.

Chapter 4: In this chapter we present an ontology-based approach to assist different financial institutions in standardizing, monitoring, and exchanging data in the credit risk field in accordance with the BCBS recommendations. We define the credit risk scorecard and delineate the major methods used for developing credit risk scorecards in addition to some examples of scorecard ontologies.

Chapter 5: This chapter suggests improving the work developed in the previous chapter. It demonstrates the implementation of credit risk scorecard and applicant ontologies based on ODPs, and complying with BCBS 239 as a stepping stone towards a more complex and sophisticated framework for credit risk management and decision making processes in financial institutions.

Conclusion and future work: This section summarizes the research and concludes with its major achievements and possible directions that could be considered for future research.

Chapter 2: Ontology and Ontologies

"Appearing to behave intelligently is not the same as behaving intelligently", Nino B.Cocchiarella

Chapter 2 Ontology and Ontologies

2. 1.Introduction

In the last decades, a huge amount of information is being produced and reproduced every day as a result of the emergence of new technologies such as social media, IoT, e-learning and e-commerce. At the same time, machines exceed humans in data storing, retrieving, integrating, and reasoning.

Therefore, to make use of this information, it should be integrated in databases which can be accessible, shared, and re-used among different scientific communities. It should also be formatted in such a way as to be exploited by machines. This objective has not yet been attained despite several attempts by researchers in sciences and information-based technologies for many reasons [12]:

1) Human and technical obstacles:
   a. Each community has its own terminology and format for describing data, or what is termed “human idiosyncrasy”.
   b. Each community uses a variety of technologies to encode and store data (technologies themselves are rapidly evolving), known as “technological idiosyncrasy”.
2) Machine limitation problems:
   a. Computers are good at computing, but not quite good at cognitive processing, so computers cannot distinguish valuable data from the rest.
   b. Two databases that store the information of disparate systems within the same or related domain cannot align their information without further intervention by human beings, if they use different terminologies or different organizational principles.

In order to resolve these problems, ontologies are a natural fit for data storing, retrieving, integrating, and reasoning. The use of ontologies to solve the above problems is very promising in this regard. It will enable us to easily process the prolific data constantly generated for the improvement of daily life and the greater benefit of humanity. In this chapter, we attempt to define the concept of ontology and its raison d’être. In addition, we tackle, in as precise and concise a manner as possible, the role of ontologies in the main fields of information knowledge systems, although there are other fields which deploy ontologies.

2. 2.Ontologies

2.2.1.Working Definition

The term ontology is borrowed from the field of philosophy; it is defined as "a science or study of being: specifically, a branch of metaphysics relating to the nature and relations of being; a particular system according to which problems of the nature of being are investigated" [13]. However, in computer and information science there are many definitions about what an ontology is. The ontology definition has changed and evolved over time. It is often context-based and molded according to objectives. For instance, the work developed in [12] is a case in point.

According to [14], one of the first definitions of ontology in computer and information science was proposed by the authors in [15]: "An ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary". Gruber proposed and developed in [16] one of the most used definitions and top cited in Google Scholar: "An Ontology is explicit specification of a shared conceptualization in terms of concepts (i.e., classes), properties and relations". The term "explicit" means it is accepted by all community members, and "Conceptualization" means an abstract, simplified view of the world that we wish to represent for some purpose. An ontology is defined by [17] as a "formal specification of a shared conceptualization.", while [18] merges the previous definitions [16] and [17] and defines it as "a formal, explicit specification of a shared conceptualization in terms of concepts (i.e., classes), properties and relations". The term " formal" entails that ontologies should be expressed in a machine-readable format such as semantic language (RDF, OWL, PLIB...) allowing automatic reasoning, consistency-checking capability, to be shared by a particular community, and able to be referenced from any environment [19].

An ontology is defined by [12] as "a representation artifact, comprising a taxonomy as proper part, whose representations are intended to designate some combination of universals, defined classes, and certain relations between them".

Taking into consideration the definitions mentioned above, this research adopts as a working definition the definition proposed by [18] since it serves the purposes of this work.

2.2.2.The Components of Ontologies

Knowledge in ontology is formalized using five kinds of components [9]:

1) Concepts: "A concept can be anything about which something is said and, therefore, could also be the description of a task, function, action, strategy, reasoning process, etc."
2) Relations "represent a type of interaction between concepts of the domain. They are formally defined as any subset of a product of n sets, that is: R: C1 x C2 x ... x Cn. Examples of binary relations include: subclass-of and connected-to".
3) Functions "are a special case of relations in which the n^{th} element of the relationship is unique for the n-1 preceding elements. Formally, functions are defined as: F: C1 x C2 x ... x Cn-1 Cn. Examples of functions are Mother-of and Price-of-a-used-car that calculates the price of a second-hand car depending on the car-model, manufacturing date and number of kilometers".
4) Instances: an element of the modeled area is represented as an instance of a class defined in ontology;
5) The axioms are predicates applied on relations, concepts, or instances; they help to ensure the integrity of data or make inferences.

Once the components of ontologies have been represented, the ontology can be encoded in various ontology languages, namely RDF, RDFS, DAML+OIL, and OWL.

2.2.3.Types of Ontologies

Many researchers are interested in the classification of ontologies such as [17], [20], and [18]; while the work of [21] depicts the existing classifications of ontologies in software systems in general, and further enriching these classifications in a manner that is tailored to ontology-enhanced user interfaces. These classifications have been established based on different dimensions. Therefore, for the sake of our research project we adopt the classification proposed and developed by the author in [20], which is highly adopted by the dominant experts in the field. He believes that the distinction among different ontologies must be done only according to their generality level and, hence, there is no reason to make a distinction among ontologies on the basis of ''the amount and type of structure of their conceptualization'', as proposed in [22] and [23].

Ontologies can be divided into three types: (i) application or local ontologies, (ii) domain ontologies, and (iii) upper ontologies (Figure 2.1).

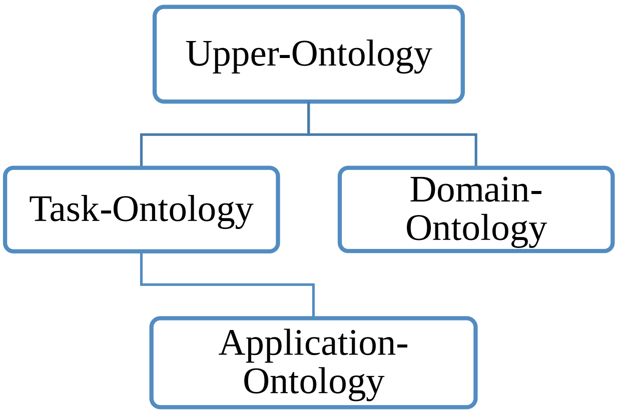

Figure 2.1. *Ontology classification according to the generality level* [20].

- Application or local ontologies: a local ontology is a specialization of domain ontology and task ontology, in order to fulfil specific needs [24].
- Domain ontologies: A domain ontology is a representation of concepts and relations that are necessary for the description of certain domains (e.g. medicine, economy, military) [19].
- Task ontologies: A task-ontology is a representation of set of primitives designed to achieve a task.
- Upper ontologies/Foundational ontologies/Top level ontologies: A top-level ontology is a high-level and domain-independent ontology whose concepts and relations are intended to be basic and universal to ensure generality and expressiveness for various domains. They permit the alignment/matching of specific domain ontologies that specialize the same upper-ontology [12], which will facilitate the data integration and knowledge reuse. Several upper-ontologies have been developed and proposed including Suggested Upper Merged Ontology (SUMO)[6] [25], Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE)[7] [26] and Basic Formal Ontology (BFO)[8] [27].

[6]See http://www.adampease.org/OP/ for more details.

[7]See http://www.loa-cnr.it/DOLCE.html for more details.

[8]See http://www.ifomis.org/bfo for more details.

2.2.4. Why Are Ontologies Important?

Ontologies are designed for the purpose of simulating the human mind and self-consciousness. Most of time, the AI community seems to be excited about some mechanisms such as IoT, Machine learning (ML), and Neural Network (NN). These mechanisms are proposed as principal ingredient for making electronic systems intelligent (e.g., computers and other AI systems). The marketplace is full of products of AI systems: manufacturing robots, disease mapping, virtual travel booking agents, and Inter-team chat tools that appear to be smart. Appearing to behave intelligently is not the same as behaving intelligently [28]. Therefore, the question is, can AI systems be intelligent? In other words, can they ever think about and understand or know what is going on? If so, how can AI systems think and be self-conscious, and what is needed to duplicate the mental states and processes of humans? The author in [28] affirms that, according to Functionalism[9], AI systems may be able to think and be self-conscious by means of AI learning programs; in any case, this will certainly require the coding of representational systems comparable to natural language. The authors in [29] confirm that ontologies are useful in natural language understanding in two ways. First, domain knowledge often plays a crucial role in disambiguation. A well-designed domain ontology provides the basis for domain knowledge representation. Second, the ontology of a domain "helps to identify the semantic categories that are involved in understanding discourse in that domain". Moreover, [30] anticipated that ontologies, more specifically formal ontologies such as BFO, will continue to be an important contribution to AI research in the years to come.

Ontology clarifies the structure of knowledge. The use of powerful mechanism theories, such as Machine Learning, in a given domain like credit has shown great potential for improving the efficiency, speed, and accuracy of tasks performed by credit analysts. The result is the models of predictions based on knowledge base; that is, without ontologies, or the conceptualization that underlies knowledge, there cannot be a vocabulary for representing knowledge [29]. **Weakness in knowledge representation systems leads to weak models.**

Ontologies enable and facilitate knowledge sharing. While academic and industry fields generate knowledge, they may not often effectively reuse and share that knowledge. Researchers and experts working in the same fields sometimes reinvent the wheels to recreate

[9]Functionalism is a theory that claims that an artificial intelligent system can in principle think and be self-conscious, and perhaps even in the same way that humans can think and be self-conscious.

the same or similar type of knowledge and intellectual work. A software engineering methodology based on formal specifications of shared resources, reusable components, and standard services is needed. Ontology is a system of vocabulary for describing the problem solving structure of all the existing tasks domains independently. It is obtained by analyzing problem solving processes of domain experts and task structures of real-world problems. Ontologies aim at capturing domain knowledge in a generic way and provide a commonly agreed understanding of a domain, which may be reused and shared with others who have similar needs for knowledge representation in that domain [29], thereby, saving time and costs, minimize risks, and increase effectiveness.

2. 3.The Role of Ontologies

In the recent decades, ontologies have been used in various contexts and domains, which explains the difficulty to identify and enumerate the different roles of ontologies in knowledge information systems. Nonetheless, several authors have focused on one or more roles of ontologies in a specific context. For our need in this work- we deal with this in a more practical way in the chapters 4 and 5 and we develop concretely the credit risk scorecard as a case study in compliance with the BCBS 239 principles- here are some prominent examples of the roles that ontologies can play in knowledge-based systems (see Figure 2.2 and Table 2.1):

- Knowledge representation and reasoning (KR&R).
- Knowledge sharing and reusing (KS&R).
- Knowledge integration (KI).
- Information retrieval (IR).
- Decision support systems (DSSs).
- User interfaces (UIs).

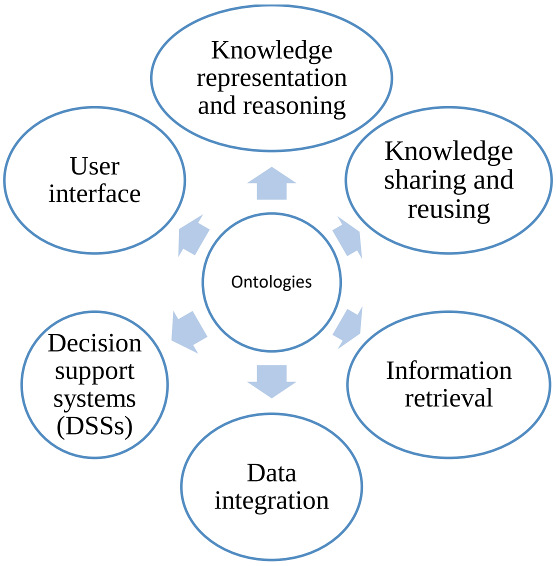

Figure 2.2. *The role of ontologies in different contexts.*

2.3.1.Ontologies and Knowledge Representation and Reasoning

Many ontologies have been developed for representing knowledge management in different contexts. These formal representations express knowledge in logical forms. With the arrival of computers and the interest of nonmathematical knowledge in AI, the AI community works tirelessly to harness formalizing declarative knowledge approaches, being developed by classical mathematical logic for the development of knowledge representation research (conceptual models and formal languages).

2.3.1.1.Knowledge Representation and Reasoning

One of the most pertinent questions that pertain to knowledge representation and reasoning is, first, what is knowledge? From the Greek philosophers to present experts in knowledge management many definitions and theories of the term knowledge were proposed; however, it is still not totally demystified and not a single one has been widely agreed upon. Contrariwise, all of them accept that knowledge 'is a justified true belief' and each one explains the ways to find the truth or justify the true belief [31].

Knowledge representation means that "the field of study concerned with using formal symbols to represent a collection of propositions believed by some putative agent" [32]. In the area of AI knowledge representation can be described as formal reconstruction of knowledge and its implementation inside intelligent system such as computers. This representation will facilitate the automatic use of this knowledge, its comprehension, and reasoning by machines.

Many specific models have been developed based on specific needs of application fields such as:

- Description logics (DLs) [33]: knowledge representation system based on DLs consists of two components (i) TBox describes domain of interest (i.e. Ontology) and (ii) ABox contains assertions about individuals in ontology context.
- Semantic networks (SNs) [34]: graph structures for representing knowledge in patterns of interconnected nodes and arcs (i.e. WordNet[10]).
- Conceptual graphs (CGs): knowledge representation language based on the existential graphs of Charles Sanders Peirce (EG)[11] and the SNs developed by [35].
- Frames: “A frame is a data-structure for representing a stereotyped situation, like being in a certain kind of living room, or going to a child's birthday party. Attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed” [36].

The authors in [37] declare that ontology-based model is one of the appropriate methods for knowledge representation approach.

The authors in [32] define the term reasoning as “the formal manipulation of the symbols representing a collection of believed propositions to produce representations of new ones.” Once knowledge is modeled and represented in knowledge base, it is ready to be used in reasoning process, a process of constructing new facts that are not explicitly expressed in the initial knowledge base. Different reasoning mechanisms are used in AI depending on the objectives of the system to be implemented, such as logical reasoning, classification reasoning, filtering, inheritance and rule-based reasoning [38].

2.3.1.2.The Role of Ontologies in Knowledge Representation and Reasoning

Ontologies constitute a great approach for the contextualization and organization of knowledge, especially in critical context such as financial analysis and risk management. In order to enrich and facilitate the re-use and maintaining of knowledge management systems,

[10]See http://wordnet.princeton.edu for more details.

[11]The system of Existential Graphs is defined as “a diagrammatic system of logic by means of which we can express, and then examine and experiment with, statements and inferences”. Don D. Roberts, The Existential Graphs, 1992.

the concepts of ontologies add semantics to knowledge resources via two different techniques: indexing and annotation. The works in [39], [40], and [41] affirm the added value of the use of ontologies as indexing and annotation resources for better understanding of the domain and knowledge re-use. Ontologies have been widely used to represent knowledge in many domains. The authors in [42] conducted a survey of biological ontologies and concluded that ontologies ensure common vocabulary for description, sharing, and querying of databases. The work in [43] developed a knowledge management system-based ontology for ICT companies whose main feature is an ontological knowledge representation, in order to improve documents' searching, sharing, and discovering. To improve the description of data and services, the work in [44] uses ontologies to add semantic information, allowing an automated processing of information and services between involved public administrations.

The knowledge representation-based ontologies can be accompanied by reasoning mechanisms. The work in [45] argues that by introducing reasoning mechanisms to a given ontology-based system, querying is achieved with more precision. Furthermore, proper performance and scalability are ensured.

2.3.2.The Role of Ontologies in Knowledge Sharing and Reusing

In order to minimize efforts, reduce costs, and save time of the development of knowledge base systems, the computer science community is paying increasing attention to the identification of strategies that can facilitate the sharing and reuse of knowledge-components, especially the ARPA Knowledge Sharing Effort [15]. They provide a new way of building an intelligent system by proposing the following:

> "Building knowledge-based systems today usually entails constructing new knowledge bases from scratch. It could instead be done by assembling reusable components. System developers would then only need to worry about creating the specialized knowledge and reasoners new to the specific task of their system. This new system would interoperate with existing systems, using them to perform some of its reasoning. In this way, declarative knowledge, problem-solving techniques and reasoning services would all be shared among systems. This approach would facilitate building bigger and better systems cheaply [15]."

Ever since, ontologies have been proposed as specific mechanisms to enhance sharing and reusing across different applications. The work in [46] developed an ontology-based system for sharing information among users in virtual communities of practice, while the work in [47] proposes domain ontologies to enrich traditional knowledge representation, providing a baseline for facilitating knowledge interpretation and sharing between humans and machines. The authors in [48] propose the use of ontologies in the product development process in two stages: (i) ontological model of product knowledge integration; (ii) exchange of product knowledge between multi-agent systems based knowledge-sharing model.

2.3.3.The Role of Ontologies in Data Integration

Data integration is the collection of data from different and heterogeneous sources, merged and fused to be processed, so that it forms a unified new whole and gives the user the illusion of interaction with one single database. Data integration can help an organization leverage information that would otherwise still be hidden. Doing so can enhance communication among organizations, as well as departments within the same organization, provide better customer service, improve decision making, and overall increase productivity.

According to the authors in [49] there are many concrete data integration solutions based on the architecture levels: (i) Mediated query systems, (ii) Portals, (iii) Data warehouse, (iv) Operational data stores, (v) federated database systems (FDBMS), (vi) Workflow management systems (WFMS), (vii) Integration by web services, (viii) Model management, (ix) Peer-to-peer (P2P), (x) Grid data integration, (xi) Personal data integration, (xii) collaborative integration and (xiii) Dataspace systems.

Several other solutions have been used for data integration systems. The work [50] affirms that ontology-based data integration systems emerged as a suitable approach for this purpose because they provide an explicit and machine-understandable conceptualization of a domain [16]. Moreover, ontologies can be used as a global query model or for the verification of the user–defined or system-generated integration description [51]. They have been used in one of the three following methods: (i) Single ontology approach, (ii) multiple ontology approach, and (iii) hybrid ontology approach as explained and detailed in [52]. Numerous researchers have developed ontology–based integration systems in different fields. The work [53] explains the coordination and the evolution of a family of ontologies for biomedical and clinical data used to annotate the multiple bodies of data, thereby making it accessible and liable to algorithmic processing. The work [54] has constructed a Dataspace based on metadata and ontology for data sharing and integration in complicated scientific research projects such as gene and proteomics. In this regard, it provides scientific researchers with a unified logical view off all data from sources in different locations and platforms; whereas the work [55] proposes an ontology as mediated schema for the representation, collection, integration, and storing of web analytics data from many sources of popular and commercial digital footprints, which can be easily queried by high level algorithms. The authors in [56] propose an ontology learning approach to integrate data and knowledge of a different nature, origin, format, and standard used for evaluating the knowledge correspondence with the learning goal, tested on English educational texts. The work in [57] brings about a new promising alternative paradigm for semantic data integration, well-known Ontology-Based Data Access (OBDA), while [58] applies the latter approach in combination with Ontology-Based Data Integration (OBDI) in social sciences and humanities. Thereby, in addition to easing data integration, this led to producing an easily read, formal specification in the form of ORM conceptual schema, allowing easy communication between different project partners, even those without computer science background.

2.3.4.The Role of Ontologies in Information Retrieval

Information Retrieval (IR) is “finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers)” [59].

In this context, the notion of ‘document’ has evolved with time. It was initially identified with text whose content was represented by keywords. Later, this notion extended to the original texts. Now, it concerns all kinds of information such as: unstructured texts, images, audios, and videos. Nowadays, millions of people engage in information retrieval every minute using different tools such as web search engines and personal virtual assistant (i.e. Apple’s Siri and Samsung’s Bixby). Hence, there is a need for the development of automated Information Retrieval Systems (IRS) capable of managing the increasing larger quantities of information distributed across heterogeneous sources, and finding precisely what people are searching for; such systems also have to be adapted to the hugely varying circumstances of their use. Ontology is such an approach that can help to develop such systems [60].

Many works have adopted the ontology approach for developing IRS for different purposes; for example, the authors in [61] developed ontology-based IRS for improving image retrieval by the users of the web. While the work in [62] build ontology-based IRS to help end users to overcome difficulties resulting from complex database search requests to retrieve information and semantic relationships between data stored in databases; therefore, extracting the result sets closer to their research requirements.

2.3.5.The Role of Ontologies in Decision Support Systems (DSSs)

The modern way to build systems is to separate data and the operational from the informational or analytical processing. Here arise data representation systems, data integration systems, and decision support systems (DSSs). DSS constitutes an interactive computer-based system that aids users in decision making.

DSSs have become more and more popular in different domains, including finance, medicine, urban planning, and military. They are especially valuable in decision making situations in which the necessary information is located in various sources and concerns different domains and in which the precision and optimality are important (e.g., in case of a pandemic as we are currently witnessing with Coronavirus outbreak). Therefore, to develop such systems, ontology is a suitable approach (i.e. as developed in [56] and [57]); a case in point is [63] which explains the approach to build intelligent DSSs using ontologies.

2.3.6.The Role of Ontologies for User Interfaces (UIs)

While a lot of systems developed and proposed have employed ontologies approach in at least one field including: Knowledge representation and reasoning, knowledge sharing and reusing, knowledge integration and information retrieval, the use of ontologies to enhance the user interfaces[12] has been gaining attention in the last years. The work realized in [21] has dealt with how the use of ontologies can enhance the user interfaces by improving:

- The visualization capabilities.
- The interaction possibilities.
- The development process of a user interface.

Also affirm that the employment of ontologies in user interfaces can add interesting features and possibilities to future software.

[12]An ontology-enhanced user interface "is a user interface whose visualization capabilities, interaction possibilities, or development process are enabled or (at least) improved by the employment of one or more ontologies" [21]

During the past years many works have used the ontology approach to improve the user interface's design; the authors in [64] use ontologies to adapt user interface according to their profile, platform, and environment allowing the adaptability of user interface in various mobile devices. While the work [65] tries to improve the personification of mobile interface for people with special needs based on the ontological representation of the interface patterns and knowledge about users and their interaction with a mobile application.

Table 2.1.*The roles of ontologies.*

Roles	Context/Problematic	Objectives	Authors
Knowledge representation and reasoning	Complexity of making inferences in molecular biology data stored in databases	Enriching and allowing common access to bioinformatics and molecular biology information	[42]
	ICT companies	Improving documents' searching, sharing and discovering	[43]
	Offering information and electronic services to citizens and enterprises	Automated processing of information and services among public administrations	[44]
	Reasoning mechanism in ontology	Achieving querying with more precision and ensuring proper performance and scalability	[45]
Knowledge sharing and reusing	Finding, organizing, and accessing information by users	Automatic information sharing in virtual communities	[46]
	Exchanging and integrating information between actors	Better exploitation of knowledge within organizations and projects	[47]
	Retrieval and reuse of heterogeneous knowledge	Exchange of product knowledge between the global and local ontologies, and among the local ontologies	[48]
Data integration	Integration of heterogeneous knowledge	Access to heterogeneous knowledge bases	[48]

Roles	Context/Problematic	Objectives	Authors
	Biological and clinical data	Making data available to search and to algorithmic processing	[53]
	Complicated scientific research data projects	Providing scientific researchers with a unified logical view off all data sources in different location and platforms	[54]
	Web analytics	Providing e-commerce SMEs with accessible tools for high level web analytics	[55]
	Composing data and knowledge of different nature, origin, format and standard	Evaluating the knowledge correspondence with the learning objective	[56]
	Social sciences and humanities data integration	Easy access of scholars to historical and cultural data about food production and commercial trade system during the Roman Empire	[58]
Information retrieval	Information retrieval from web pages	Improving image retrieval in the web	[61]
	Complex database search requests to retrieve information	Aiding end users to retrieve information	[62]
Decision support systems (DSS)	Web analytics	Performing customers' profile analyses using advanced data mining algorithms	[55]
	Knowledge discovery means	Evaluating information pertinence	[56]
User interfaces	User Interface in a variety of mobile devices	Adapting user interface according to user's profile, platform, and environment	[64]
	User mobile interface for people with special needs	Personification of mobile interface for people with special needs	[65]

In this section, we have tackled, in as precise and concise a manner as possible, the role of ontologies in the main fields of information systems, although there are other fields which deploy ontologies. The rationale behind this section, however, is not to provide a conclusive view or overview of the use of ontologies as much as it is to provide a clear idea about the main uses of ontologies and their utility, rather necessity, in information systems domains. There is no doubt that the best approach to implement the BCBS 239 principles, facilitated by the best practices proposed by [66], is the use of ontologies. This approach heralds promising results in terms of reusability and expendability of ontologies as well as saving time and reducing the cost of implementing the solutions suggested- the cost of such solutions as estimated by the McKinsey 2015 report is approximately 230 million USD per an average globally systemically important bank (G-SIB) and 75 million USD per an average domestic systemically important bank (D-SIB). Furthermore, the use of ontologies will bridge the gap between theory and practice and help to solve the problem of further delay in complying with the BCBS 239 principles; banks failed to meet the January 1, 2016 deadline issued by the BCBS self-assessment exercise to 31 G-SIBs in 2014 [67]. It is therefore high time we invest in ontologies and put them to use for all the benefits we have shown above. It is also noteworthy that the AI community is not ignorant of this; they have realized that ontologies will play a primordial role in developing intelligent machines and revolutionize the way we deal with data, and hence ushering a new era of wisdom in dealing with data.

2. 4.Conclusion

In this chapter, we attempted to highlight the added value of ontologies in overcoming the current setbacks which lie in human and technical obstacles, on one hand, and machine limitations problems, on the other, which stand against taking best advantage of the huge amount information available. In this regard, we have sought to reveal the primordial role of ontologies for the sake of this study. Thereby, we reviewed a set of ontology definitions, which have evolved over time, taking into consideration the problematic aspects of defining an ontology; that is, it is often context-based and is crafted as the objectives demand. Ensuing this discussion, a brief overview distinguishing among different types of ontologies is also provided to draw a bigger picture about ontologies. The last part of this chapter deals with the use and role of ontologies in improving each of the following information systems fields: knowledge representation and reasoning, knowledge sharing and reusing, knowledge integration, information retrieval, support decision systems, and user interfaces. It goes without saying that these fields are by no means conclusive, for that is not the purpose of this work. The main objective, however, is to show the ontology's great potential in addressing problems related to data storing, retrieving, integrating, and reasoning, as well as to highlight its contribution to AI research, given its promising role in the future. In short, ontologies are no longer a luxury or an option among many other options we can pick up from; rather, they are a necessity which is asserting itself day by day. Ontologies will not only help us store, retrieve, integrate, and reason on data, but they will also save our time and costs, minimize our risks, and increase our effectiveness.

Chapter 3: Ontology Engineering with Ontology Design Patterns

"The semantic web is a set of standards and best practices for sharing data and the semantics of that data over the Web for use by applications", Bob DuCharme.

Chapter 3 Ontology Engineering with Ontology Design Patterns

3. 1.Introduction

In the last decades, the use of ontologies has become more and more popular in both academia and industry fields ranging from e-learning to financial, medicine, military and engineering systems. However, many ontology projects have failed due to, at least in part, a lack of discipline in the development process; that is, the poorly specified, underspecified, or lack of requirements and evaluation criteria [68]. Developing a successful ontology depends on how far its methodology supports collaborative construction and allows for reusability. The methodology should also take into consideration the degree of application dependency, life cycle recommendation, strategy for identifying concepts, methodology details, and interoperability support.

Therefore, ontologies shall be well developed to fit purpose and to be re-used. In this chapter we aim to focus on the ontology engineering with ontology design patterns.

3. 2.Ontology Development Methods Review

The development of ontology is an important step in the development of knowledge base systems; the advantages of such system-based ontologies have been broadly discussed and presented in chapter 2. However, the development of ontology is a tedious and time consuming task. It requires significant and qualified skills as an art and creativity rather than technology or understood engineering process. Although a variety of ontology building methodologies have been proposed, they are still at the level of general guidelines and none of them are mature enough to be standardized or receive wide consensus in the ontology community [69], [70].

Generally speaking, most of these methodologies are described as an iterative ontology engineering process; i.e., the deliverable of each step of the process can be modified, revised, evaluated, and refined at any stage of the process. This process may include, albeit with different naming, the following steps: identifying the ontology scope and the ontology capture, choosing the ontology encoding language, integrating the ontology with the existing one, evaluating the developed ontology, and documenting the ontology [71]. They also follow the same scenarios depending on the knowledge resources available: without reusing existing resource (i.e., the ontology developed from scratch), non-ontological resources or ontological resources, e. g., ontology design patterns reuse as is the case in this work (see Figure 3.1).

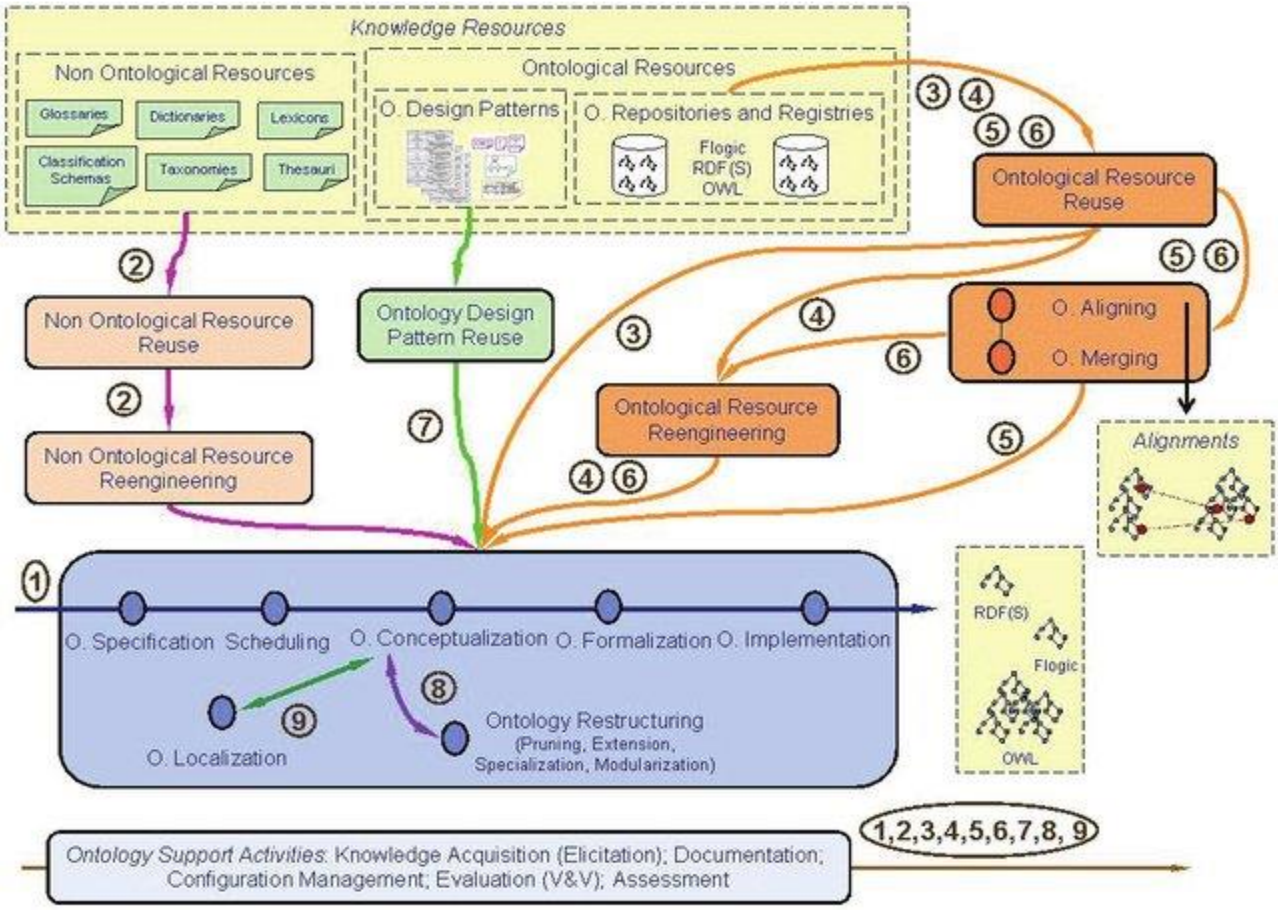

Figure 3.1. *Scenarios for Building Ontology Networks* [72].

3.2.1.Methods and Methodologies for Building Ontologies

There are many methods inspired by software development methodologies proposed for the development and maintenance of ontologies. Some of these methodologies are designed for building ontologies from scratch, reusing non-ontological resources, or ontological resources. In this section, based on the same evaluation criteria used by Iqbal et al. [73], we provide a review of the most popular methodologies used by developers in order to come up with the most suitable approach for our work (see Table 3.1). The choice of adopting these criteria is owed to the fact that the study, apart from being the most cited one in Google Scholar, is as conclusive as possible, which allows for greater flexibility in the evaluation and choosing of the best methodology. It is also noteworthy that we have added five more methodologies (most of them came out after the study) to the two borrowed from the study (the most prominent ones namely METHONTOLOGY and Gruninger and Fox) in order to compare them before finally selecting our approach.

3.2.1.1.Gruninger and Fox Methodology

Gruninger and Fox methodology [74] is developed based on experiences in the development of TOVE (Toronto Virtual Enterprise) project. It is a semi-independent application [75] inspired by the development of knowledge-based systems using first order logic. It proposes six steps for building an ontology:

1) Identify motivating scenarios: the starting point for ontology development is a set of scenarios which are story problems or examples that cannot be solved or adequately addressed by existing ontologies. A set of solutions to the scenario problems may include motivating scenarios; these solutions provide an informal intended semantics for objects and relationships between those objects that will be included in the ontology.
2) Formulate informal competency questions (CQs): given a set of scenarios obtained in the first step, an ensemble of CQs are formulated to provide an initial evaluation of the new or extended ontology. These CQs are in natural language forms until they are expressed in a given formal language of ontology.
3) Specification in First-Order Logic – Terminology: once the informal CQs have been posed, a set of terms is extracted to serve as a basis for specification of the ontology in a formal language. Then, the terminology is specified using first-order logic (formal language such as OWL or similar).
4) Formal competency questions: the informal competency questions are formalized using the same formal language as the one used for the terminology.
5) Specification in First-Order Logic – Axioms: define the axioms using first-order logic guided by the formal competency questions in iterative process (i.e., update the set of axioms added to the ontology until it is sufficient to express the formal competency questions). These axioms will be used to determine the semantics or meaning of ontology’s terms.
6) Completeness theorems: the last step is the evaluation stage; once the competency questions have been formally stated, the conditions under which the solutions to the questions are completed must be defined.

3.2.1.2.METHONTOLOGY Methodology

Based on the experience acquired in the development of ontology in chemical domain, the Ontology Engineering Group at Universidad Politécnica de Madrid developed the METHONTOLOGY framework [76] which is used to build ontologies from knowledge level. It proposes two levels of ontology construction: (i) an ontology development process and (ii) an ontology life cycle. The ontology development process identifies a set of activities to be carried out to develop ontologies: planning tasks, specifying purpose and scope, acquiring knowledge using Knowledge Base Systems (KBSs), conceptualizing this knowledge in a conceptual model, formalizing this model in order to turn it into a formal or semi-compatible model, integrating existing ontologies, implementing the ontology in a formal language, evaluating it, documenting it, and carefully maintaining it. The ontology life cycle allows the ontologist to revise and modify the deliverable of any development process steps: specification, conceptualization, formalization, integration, implementation, and maintenance, regardless of the current stage of development process. However, regardless of its main limitations, the methodology enables the construction of ontologies at knowledge level and does not take in consideration the scenarios of the reuse and reengineering of non-ontological resources and the ontological resources.

3.2.1.3.NeOn methodology

Contrariwise to previous methods that help develop ontologies from scratch, the NeOn methodology [72] has initiated a new paradigm whose emphasis is the reuse and possible subsequent reengineering of knowledge aware resources. It proposes nine flexible ontology-building scenarios- these scenarios can be combined in different ways- (see Figure 3.1). To build ontologies and ontology networks, there are three activities and processes to be carried out during the whole ontology and ontology network development: (i) knowledge acquisition and elicitation, (ii) configuration management, and (iii) evaluation and assessment.

The methodology supports the argumentation process (the collaboration is mentioned but not treated in detail) which permits the collaborators to reach a consensus or explain a decision about the requirements of the ontology networks and how they should be implemented during different scenarios of building ontologies. Each discussion or decision during the entire ontology building process can be documented or even recorded. The methodology proposes a set of guidelines and tasks to carry out the ontology specification activity:

- Identifying purpose, scope, and level of formality: obtaining the main goal, scope, and granularity of ontology.
- Identifying intended users: determining the main users of the ontology.
- Identifying intended uses: obtaining a list of intended uses via a set of scenarios.
- Identifying requirements: identifying a set of requirements or needs that an ontology should fulfill once formally implemented using CQs, tools as mind map tools, excel, and collaborative tools techniques.
- Group requirements: a set of groups including different CQs that are relevant to specific features of the ontology.
- Validating the set of requirements: identifying possible disagreements between CQs, missing CQs, and contradictions in CQs.
- Prioritizing requirements: determining a set of priorities related to each group of CQs and for each CQ in the group.
- Extracting terminology and its frequency: extracting the terms that will be formally represented in the ontology.

3.2.1.4.Basic Formal Ontology Methodology

The BFO methodology [12] is a top-down approach to the problem of automatically administrating scientific information located in heterogeneous data sources. To ensure widespread accessibility and usability, it proposes to start from defining the general scientific concepts then encoding detailed terminological content of a specific science. To achieve this goal, the methodology proposes the following steps:

In order to design a useful ontology, the methodology proposes to keep in mind and to respect four general principles:

1) Realism: the idea behind the development of ontology is to represent reality, not people’s concepts or mental representations or use of languages. Science can help to determine general features of reality in the form of terms and relation between them.
2) Perspectivalism: it comes from the postulate that reality is too complex and variegated. It needs two or more scientific theories to be covered in its totality. The ontology developers should respect a modular approach rather than seek to represent all portions and features of reality in one single ontology. The ontology developers can

develop multi-ontologies to accurate descriptions of reality in which each one should be maintained by experts in the corresponding scientific discipline.

3) Fallibilism: it flows from the fact that scientific theories may be subject to correction. Therefore, the implication of fallibilism on ontology design is that every ontology:
    - Must have strategies of versioning.
    - Needs to have tracking service for it users.
4) Adequatism: given the fact that each scientific discipline provides a representation of what exists in reality. Hence, the ontology in a given domain should be designed to represent the entities, not kinds of entities.

The methodology proposes to take into consideration other four concrete guidelines concerning the design process itself:

1) The principle of reuse: the first step in ontology development should be searching, examining, and evaluating existing ontological resources in the domain and around for possible reutilization or at least for recommendation.
2) The ontology design process should balance utility and realism: the ontology should be designed to meet specific local purposes provided that the ontology design process should not be on the detriment of adequacy to the reality which the ontology is being developed to represent. The ontology should be developed to be re-used in neighboring domains to which an ontology is constructed for.
3) The ontology design process is Open-Ended: scientific ontologies should be designed in such a way as to be expandable and adjust to neighboring ontologies through time; i.e., they should be in continuous maintaining, evaluating, updating, correcting, and adjusting.
4) The principle of Low-Hanging fruit: the ontology designer should start by identifying the general terms most commonly used that are easier to understand and define, then move step by step to more complex and controversial terms.

The BFO methodology is a top-down approach. In the first stage, to design a good ontology it proposes, as previously outlined, to follow general principles in order to determine general features of the ontology. In the second stage, it proposes a set of steps to follow for the development of a domain ontology:

1) Demarcate the subject matter of the ontology: determine the intended scope of the ontology.

2) Gather information: select the most common general terms from relevant ontologies, such as BFO, and from standard textbooks.
3) Order these terms hierarchically from the more to the less general ones.
4) Regiment the results: proceed to an iterative process to understand well the domain, starting from the deliverable of phase 3 in order to guarantee:
    - The hierarchy coherence which will serve as a core to the ontology being developed.
    - The human understandable definitions for the selected terms.
    - The coherence with neighboring ontologies.
5) Formalize the ontology: encode the ontology using the formal language via an iterative task. Once the terms have been extracted, hierarchized, and defined in natural language, they are specified using first-order logic (formal language such as OWL or similar).

3.2.1.5.Agile Methodology for Ontology Development (AMOD)

AMOD [77] is one of agile methodologies which is used in the ontology engineering. It aims to adapt the agile principles and practices from software engineering into the development of ontologies. The AMOD framework consists of the following parts:

1) The ontology development:
    - Pre-game phase includes (i) identification of ontology goal and scope, (ii) specification of the tools and techniques used for the formalization and the implementation of the ontology, (iii) identification of ontology requirements using competency questions, and (iv) selection of available sources allowing the extraction of domain knowledge.
    - Development phase is organized in multiple and iterative cycles named sprints, each sprint includes the following stages: (i) identification of items and how they are implemented during the sprint planning; (ii) capturing the terms and relations between them using some techniques such as interviewing and brainstorming (knowledge acquisition); (iii) organizing the terms in hierarchy as the core of ontology being developed; (iv) transforming the ontology conceptual model into a formal model using a formal language such as OWL or similar; (v) once the ontology is developed in this sprint, it must integrate those developed in the previous sprints; (vi) and finally reviewing the sprint by

the ontology engineer and ontology owner.

- Post-game phase includes the following activities: (i) evaluation of ontology with regard to ontology consistency, answering CQs, and ontology content; and (ii) maintenance of ontology.

2) Support activities:

Supporting activities are achieved in parallel with the development of ontology; they include the following activities:

- Ontology documentation includes three main aspects: creating a human-readable representation of the ontology content, creating machine-readable annotations of documentation metadata, and making the documentation files available as a web resource.
- Configuration management includes four activities: configuration identification, configuration control, and configuration control and configuration audits.

The work [78] presents a simplified agile methodology for ontology development in the form of quick, small, and iterative steps to produce ontologies ready to be used and easily understandable. It contains the same framework as AMOD on miniature scale which aims at developing the final model through a series of small steps. Thus, it can be used to solve and model a small and simple enterprise project in which the ontology being developed is composed by a limited amount of ontological entities.

However, the work [79] affirms that using agile in software and development of safety critical applications requires more efforts than just applying it. Therefore, the use of agile may impact negatively the product innovations.

3.2.1.6.Krisnadhi and Hitzler Methodology

The methodology proposed in [80], [81] is an approach based on a worked example for developing the so-called modular ontologies based on Ontology Design Patterns (ODPs). Developing ODPs in a particular domain typically requires domain expert team. The methodology proposes to adopt the following workflow:

1) Use case(s) or scope of use case(s): Define the use case(s) or a set of potential use case(s) for which the ODP is intended and that shall drive the modeling of ontology.

The modeling should be as general as possible which will permit to create an underlying schema which is robust, expandable, and easy to maintain and update.

2) Competency Questions and data sources: Formulate in natural language a set of CQs that shall be answered once the ontology is populated. Inventory the initial list of available data sources that can be used for the intended purpose. This list can be updated with time.
3) Key Notions to model: Build an ontology for the data we already have, lead to an ontology that is really only useful for a specific use-case, so as soon as data or data formats change, the ontology will quickly lose its value leaving no way to maintain it and to expand it or reuse it by another community or enterprise. To attempt to counteract these issues and to construct a successful ontology, the structure of the latter will not be informed only by the data or data formats but by general ontology design patters. Based on data sources and CQs, identification of the pattern which will be used is better borrowed from best practices in the library [82] in order to realize the very common ontology design pattern for each purpose. The starting point for this step is identifying the pattern which will be the core of modeling, adapting/changing it as needed, and adding axioms to each module informed by the pattern axioms, obtaining as result a set of modules for the final ontology.
4) Putting things together: Assemble the modules developed in step 3 together and check module axioms for consistency.
5) Create OWL files:

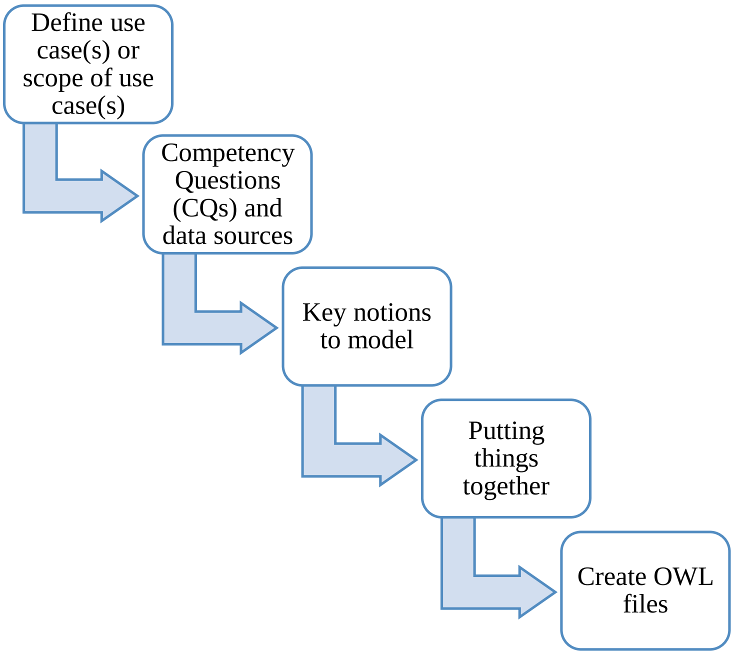

Figure 3.2. *Krisnadhi and Hitzler methodology life cycle.*

3.2.1.7.Kendall and McGuinness methodology

Based on their most cited ontology engineering method [83], according to Google Scholar, and on their experience in academic and industry fields, McGuinness and Kendall wrote the book Ontology Engineering (2018) [68] in order to demystify the holistic ontology development process. In chapter 3 of the book they deal with the first step of ontology engineering methodology; they propose a set of steps to determine requirements and use cases:

- Use cases [84], [85] are used to capture what is needed in terms of domain knowledge reference materials, ontology-specific requirements, basic functional requirement and success criteria, and to determine scoping boundaries of ontology development. The deliverable of this step is iterative and keep up to date, so it can be modified, revised, evaluated, and refined at any stage of the ontology development process. They are documented to reflect preliminary requirements, requirements learned through development, testing, deployment, maintenance, and evolution.
- Gathering resources and potentially reusable ontologies.
- Extracting domain terminology.
- Creating usage scenarios as they help in describing the use case. They provide

the frame for the set of steps that together form the normal flow of use case operations.

- Determining the flow of events: the normal or basic flow of events (also known as the primary scenario) defines the process whereby a system that is implemented is executed from start to end. Alternate flows, however, remain complementary to the basic flow. They are often invoked by valid, but not typical, situations or rules.
- Creating competency questions (CQs): an ensemble of questions that a knowledge base or application must be able to address correctly [74], [86]. CQs are derived from user stories, usage scenarios, normal and alternative process steps, as well as from reference materials.
- Describing additional resources: additional resources are composed of miscellaneous notes (information about non-human actors) collected at the very end of the use case and then moved around once we understand where they fit in the overall requirements.
- Integrating an ontology into a larger business requirements document (BRD) to provide a basic vocabulary, related definitions, and the seed for elaboration not only of the ontology itself but for other project components.

In chapter 4, they tackle the essential components of ontology engineering terminologies and their management. The "terminology work involves concepts, terms that are used to refer to those concepts in specific contexts, and formal vocabularies, including the establishing definitions for those concepts" [68]. Terminology is imperative in providing adequate common vocabularies in institutions which will facilitate sharing data within and across institutions, and enable them to measure their performance. For instance, the main problem of the 2008 financial crisis is the lack of ground-truth in many banking data assertions or the decisions that depend on the data, creating a confusion in terms used to manage risks. Though the terminology is sine qua non for the development of ontology, this methodology, however, does not expand on this part as to include proper documentation of the terminology identification.

In the final chapter of the book, they present the primary steps involved in the conceptual modeling aspects of ontology development, starting from the deliverable of the previous stages namely: one or more use case(s), CQs, list of curated terms, and a business architecture, and others business models if available. Given those input, the ontologist can formalize the ontology in formal language following a set of activities:

- Reaching existing ontological resources to determine reusability.
- Identifying relationships between the concepts from the term list and CQs.
- Testing, evaluation and validating of the ontology.

Table 3.1.*Comparison of methodologies based on the established criterion.*

Methodologies	Type of Development	Collaborative construction	Reusability support	Degree of application dependency	Life cycle recommendation	Strategies for identifying concepts	Methodology details	Interoperability support
Gruninger and Fox (1995)	Stage based model	No	Yes	Application Semi-dependent	No	Middle-out approach	some details	No
METHONTOLOGY (1997)	Evolving prototype	No	Yes	Application independent	Yes	Middle-out approach	sufficient details	No
NeoN (2008)	Stage based model	Yes	Yes	Application dependent	Yes	Middle-out approach	some details	Yes
BFO (2015)	Evolving prototype	No	Yes	Application independent	Yes	Top-down approach	Some details	Yes
AMOD (2016)	Stage based model	No	No	Application dependent	Yes	Middle-out approach	Some details	No
Krisnadhi & Hitzler (2016)	Evolving prototype	Yes	Yes	Application semi-dependent	Yes	Middle-out approach	Some details	Yes
Kendall and McGuinness (2019)	Evolving prototype	Yes	Yes	Application independent	No	Developer's consent	Some details	No

3. 3.Fundamentals

Working with Krisnadhi & Hitzler methodology needs an ensemble of technologies and tools. In this section we define and describe these fundamentals in order to account for the development process of our ontologies. The success of an ontology depends in part on such fundamentals.

3.3.1.Ontology Design Patterns (ODPs)

Ontology Design Patterns (ODPs) are reusable solutions for recurrent ontology modeling problems, which permits our ontology to be reused and expanded. [87], [88] first proposed this idea with the aim to clearly design ontologies used as the basis for building other ontologies and to foresee the effect of changes or extensions to them. It is also rarely the case that you, as an ontology engineer- having the set of requirements for your ontology engineering task at hand- will fully agree with all the ontological commitments that are made in such a large ontology. However, not reusing any well-established practices at all, and not aligning your ontology partly with existing ontologies will create problems in the interoperability, and potentially also, understandability and reusability of your ontology. It will also lead to wasting much more time and effort. Hence, there is a trade-off between interoperability, on the one hand, and over-commitment and conflicting requirements on the other hand. This is where ODPs as small "building blocks" offer one way to manage this trade-off. ODPs carry the promise of better integration and interoperability of data across various domains [89]. There are several types of ODPs and they can be reused and applied in many different ways. [90] has identified them and grouped them into six families shown in Figure 3.3: Structured OPs, Correspondence OPs, Content OPs (CPs), Reasoning OPs, Presentation OPs, and Lexico-Syntactic OPs.

A Content OP (CP) can be considered roughly as analogous to a software design pattern with the added benefit that it includes a reference base implementation (in the form of an OWL building block) ready for immediate customization [91].

In this work, we have chosen to use CPs as defined in the NeOn Project [92] to model our ontologies, as this is the most common type of ODPs with some 100+ patterns published [93].

[90] noted that each CP is associated with a catalogue entry including the following set of information fields: Name provides a name for the pattern; Intent describes the generic use case addressed by pattern; Competency Questions (CQs) contains examples of CQ that the

knowledge base associated with the CPs need to address; Diagram depicts a UML class diagram representing the pattern; Elements describes the classes and relations included in the pattern. Also, CPs are described with the Scenario, Consequence, and Known, Extracted from/Reengineered from ontological, non-ontological or related patterns.

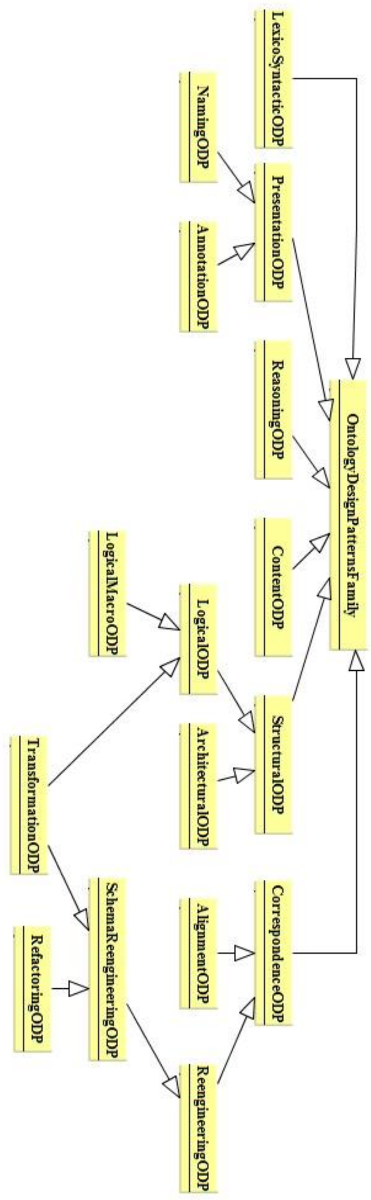

Figure 3.3. *Ontology design patterns family* [94]

3.3.2.Ontology Languages (RDF, RDFS and OWL)

Ontologies are formal theories about a certain domain; therefore, they require a formal language to encode them. This ontology language should meet the essential following requirements as proposed in [69], detailed and enriched in [95]. “It must have:

- Expressiveness/constructs/modelling features;
- Features of the language as a whole;
- Usability by computer;
- Usability & modelling by humans;
- Interaction with outside;
- Ontological decisions.”

According to [96], ontology languages are classified into three categories:

1) Traditional ontology languages: they are Description Languages (DL) based languages, languages based on first-order predicate logic, frame-based languages and others.
2) Web-based ontology languages: they are compatible with web standards category.
3) Web standards: they are used to interchange data in the internet.

A hybrid-category can be created based on the web-based ontology languages and traditional ontology languages categories such as OWL DL [97]. In this section we deal with the most relevant ontology languages currently available, namely RDF(S), OWL.

3.3.3.Resource Description Framework Schema RDF(S)

The Resource Description Framework (RDF) [98] is the World Wide Web Consortium (W3C) recommendation for semantic annotation in the semantic web. “The development of RDF is an attempt to support effective creation, exchange and use of annotations on the Web” [99]. RDF annotates the resources using the principle of universality: “any resource can be identified by a Uniform Resource Identifier (URI)" [100]. RDF uses several serialization formats, classified in two categories: (i) XML serialization formats such as RDF/XML, and (ii) non-XML serialization formats such as turtle [100], the format used in this dissertation. Each RDF statement consists of a triple: subject-predicate-object, to represent this statement in a machine-processable way. However, the RDF does not provide any mechanisms for declaring these properties nor for defining the relationships between these properties and other resources; this is where the role of RDF Schema comes. RDFS is a semantic extension of RDF and provides data-modeling vocabularies for RDF data. The combination of RDF and RDFS is commonly known as RDF(S).

3.3.4.Ontology Web Language (OWL)

The Ontology Web Language (OWL) developed by W3C is “a language for representing ontologies that is based on formal logic, a discipline that evolved from philosophy and mathematics” [101]. It is strongly inspired by the DAML+OIL language, built upon RDF(S) (see Figure 3.4) to define ontologies. It includes three sublanguages, namely, OWL Lite, OWL DL, and OWL Full, with different levels of expressivity and reasoning capabilities. OWL DL is more expressive than OWL Lite, and OWL Full is more expressive than OWL DL.

1) OWL Lite: it is the least expressive variant of OWL, mainly designated for class hierarchies; simple constraints feature in comparison to OWL DL. OWL Lite supports cardinality with only values 0 and 1 [69].
2) OWL DL: it is based on Description Logics theoretical properties; it provides maximum expressivity without losing computational and decidability of reasoning systems [102]. OWL DL support all OWL language constructs with restrictions such as type separation [102]. In addition to OWL Lite features, OWL DL has many other features including Full cardinality, Disjunction, Enumerated classes [69].
3) OWL Full: it is a complete OWL language which provides all OWL language constructs with complete and efficient reasoning. OWL Full is a unique variant of OWL which is fully compatible with RDF, both semantically and syntactically, and provides the maximum expressivity and flexibility to represent ontologies [69].

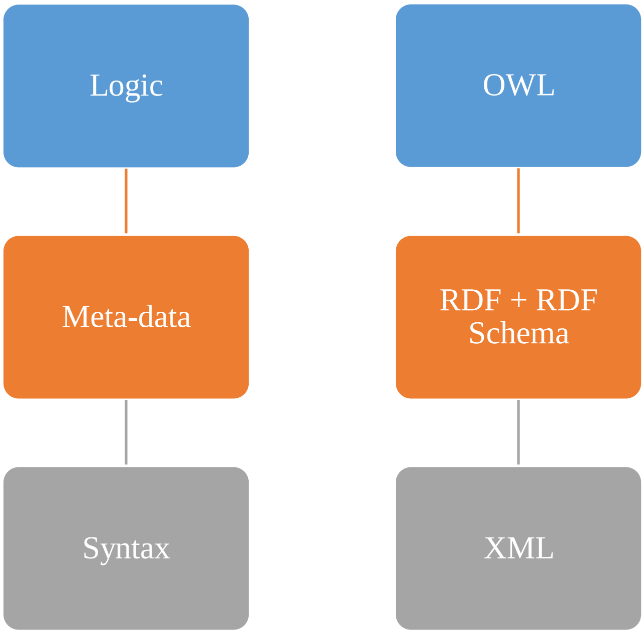

Figure 3.4. *The basic layers of semantic web*

3.3.5. Ontology Query Languages

Recently, the massive proliferation of data stored in ontology-based knowledge bases produced the need for the development of ontology query languages. The aim is to facilitate the exploitation of and access to various types of data sources by end-users and applications. For this purpose, some query languages have been developed and are still in the making including SPARQL.

SPARQL

In 2008, the W3C group recommended SPARQL as the standard query language for RDF [103]. SPARQL was released by the RDF Data access Working Group (RDF-DAWG) [3]. The part "RQL" in the acronym SPARQL means that it is designed to query and manipulate the RDF "files" stored in different formats such as XML, JSON, and Turtle, or a combination of these formats. The "protocol" part determines how communication between a client program and SPARQL server should be when they exchange the SPARQL queries and results. Therefore, as its name suggests, SPARQL query result document is a standard XML format. In this work, we have considered ARQ SPARQL processor [104], qualified as one of the best processors [105], for querying and validating our ontologies.

3.3.6. Ontology Development Tools

Enormous ontology development tools have been developed for the formalization and the implementation of the ontology using different ontology languages such as RDF, RDFS, and OWL. However, the most dominant and domain-independent tool is Protégé [106]. Protégé [107] is a free open source platform that allows users to develop reusable ontologies and to build knowledge-based systems.[13]It provides support development for ontologies in RDF(S) and OWL. It comes in two varieties: (i) Protégé desktop is a feature rich ontology editing environment enabling the construction and maintenance of OWL ontologies (Ontologies of this work have been developed in Protégé desktop 5.1.0.), and (ii) WebProtégé is a web-based version that offers distributed collaborative viewing and editing of OWL ontologies [107].

3. 4.Conclusion

The main factor for a well-developed ontology rests on the choice of the ontology engineering methodology adopted. The successful, however, of any ontology is measured by its adoption in a given community/domain. Several methodologies have been proposed by both academics and professionals since their emergence in the early nineties with Lenat & Guha, 1990. However, none of the methodologies developed so far enjoy consensus in the community. This is primarily due to the lack of some evaluation criteria or/and insufficient details regarding certain criteria. These shortcomings remain, up to this date, the main reason behind the failure of many ontology projects.

Despite all the efforts done in methodology development, ontology engineering methodologies are still a difficult process, and many challenges remain to be solved. According to our review of the most prominent ontology engineering methodologies, the most mature and suitable approach for the development of ontology, especially in a critical domain such as credit risk management, is Krisnadhi & Hitzler methodology; the latter is perceived as being more practical and helpful than others such as NeOn and METHONTOLOGY. Krisnadhi & Hitzler methodology, based on ontology design patterns, ensures that there is a trade-off between interoperability, on the one hand, and over-commitment and conflicting requirements on the other hand. This is not to say though that this methodology is perfect and with no limitations.

[13]The editors of Protégé at Stanford University highlight its importance as follows: "Protégé's plug-in architecture can be adapted to build both simple and complex ontology-based applications. Developers can integrate the output of Protégé with rule systems or other problem solvers to construct a wide range of intelligent systems."

Chapter 4: Credit Risk Scorecard and Applicant Patterns

"We must be systematic, but we should keep our systems open." Alfred North Whitehead

Chapter 4 Credit Risk Scorecard and Applicant Patterns

4. 1.Introduction

In the 2010 talk, Gov 2.0 Expo in Washington, D.C., Tim Berners-Lee suggested awarding stars for any kind of sharing data [3]. Therefore, governments earn three stars for sharing data on the web a nonproprietary format, for stars on putting it in linked data and a full five stars for connecting data to other data. Following the recommendations of BCBS and Tim Berners-Lee, an ontology-based approach has been proposed to assist different financial institutions in standardizing, monitoring, and exchanging data in the credit risk field.

The work presented in this chapter proposes an ontological scorecard model for credit risk management. It sets to define the credit risk scorecard and delineate the major methods used for developing credit risk scorecards in addition to some examples of scorecard ontologies. We conclude the chapter by presenting the development of our credit risk scorecard pattern. Krisnadhi & Hitzler methodology is used to enable the implementation of the domain knowledge and to improve the decision-making process in credit monitoring. We have developed, in this chapter, the first part of our system, which establishes a shareable ontology that can be understood both by humans and computers alike. The modeling of our ontology will make the data publication simpler and graph structures intuitive, thus, making our system open to the extent that it allows for constant accumulation of tacit knowledge and efficient polymerization of explicit knowledge.

4. 2.Credit Risk Scorecard

4.2.1.Definition

The credit-granting process leads to two choices: granting a loan to a new customer or declining their application. The purpose of the credit risk scorecard is defined as “the assignment quantitative measure to a potential borrower to provide an estimation of its capacity to repay a loan” [108]. Usually, logistic regression is the most used technique to build credit risk scorecard. More recently, artificial intelligence techniques like expert systems and neural networks have been used [109]. All of them involve establishing and quantifying the relationship between the characteristics and good/bad performance (target).

During the last few decades, credit risk scorecards (see Table 4.1) have been widely used in the banking area due to the proliferation of credit applications for different bank products, including applicants' loan [2]. The credit scoring is a statistical tool which allows banks to determine the risk level of an applicant or a customer. There are two types of credit risk scorecard models [110]: (i) Application Scoring (AS) determines the score of future credit applications and deciding which ones to accept and which to reject based on information history about applicants; (ii) Behavioral Scoring (BS) forecasts the behavior of a customer (good or bad) after the loan is made with respect to its current position. The main role of credit risk scorecard is how to distinguish between good and bad customers. By definition the applicant is good in general when he/she is credit-worthy, has no unpaid instalments or one payment in arrears; contrariwise, the applicant is bad due to the lack of repayment, and usually has three consecutive or more payments in arrears.

Credit risk scorecard uses the credit organization's historical information, namely variables such as age, gender, date of birth, employment status, etc. gathered by banks, describing the default behavior of their customers. Scorecard variables maybe selected from any resources including credit bureau data, social network, and demographic data [109]. It should include between 6 and 15 variables to account properly for the applicant's behavior [110]. Each variable is classified by an ensemble of categories (e.g. variable: "Age" and categories: > 25 years and <= 25 years). Based on statistical analysis, each category is assigned points, which will be used to calculate the total score of an applicant. Each category of each variable is assigned point which make up the total score of each applicant. Many financial institutions, especially in Europe, decided to take on analytics to improve and be more competitive. Moreover, they opted for bringing credit risk scorecard development in-house to comply with the Basel II accord [109]. This allows the banks to reduce the costs and losses, as well as be more effective.

4.2.2.Methods Used to Develop Credit Risk Scorecard

The classification technique is one of the most studied and used method to build efficient and effective credit scoring system that can be effectively used for predictive purposes. It is divided into five groups: (i) statistical models; (ii) operational research methods; (iii) artificial intelligence techniques; (iv) hybrid approaches; and (v) ensemble models. The statistical models contain several techniques- including logistic regression techniques, linear discriminant analysis, k-nearest neighbor, and classification tree. Logistic regression is widely used in industry, and especially in banking [2], [110], [111] and it is the most appropriate approach [112] for building credit scoring models due to its openness, simplicity, and ease of compliance [109]. For the purposes of this work, we work with credit risk scorecard developed using logistic regression.

Logistic regression is one of the most widely accepted statistical technique in the industry field. The work [113] claims that it was originated in the study of population growth in 1798. Logistic regression has been plenty used in financial industry applications (e.g. credit scoring application), partly because: "(i) it is specifically designed to handle a binary outcome; (ii) the final probability cannot fall outside of the range 0 to 1; and (iii) it provides a fairly robust estimate of the actual probability, given available information" [113].

Logistic regression has iterative procedures to select the final characteristics to find the best possible model based on three different types of stepwise logistic regression techniques:

1) Forward selection: it starts with zero variables, then selects the best variable based on the individual predictive power of each characteristic; same thing for the second one, and so on for the rest of variables until no remaining variable had added value on the scorecard.
2) Backward elimination: the opposite of forward selection technique; it starts with all variables in the scorecard and removes each variable that provides little value, given the other variables in the scorecard.
3) Stepwise: a combination of two precedent techniques implicates that removing and adding the variables to scorecard can be done in each step until reaching the best.

4.2.3.Credit Risk Scorecard Ontologies

The ontology was used in the financial industry; the unique and sole reference in the field is the Financial Industry Business Ontology (FIBO) [114]. It is a family of ontologies developed to standardize the terms used in the financial services industry. However, FIBO ontologies have some drawbacks such as:

1) The methodology used to build and maintain FIBO based on Agile Development Software principles and practices [115]. Initially, it was designed and adapted to support software development not artifacts such as ontologies. However, works done in this field are not mature enough to be largely approved by organizations [77]. In addition, it was identified that ontology refinement and validation phase can be developed iteratively, whereas other development phases involved larger challenges

for adopting agile methods [79].

2) The absence of clear documentation and design conventions in the FIBO standards, such the documentation of the Gene Ontology (GO) [116], and Enterprise Ontology (TOVE) [117], gives the impression that certain classes and properties may be considered to be polysemic, which makes their re-usability in different financial industry applications by end-user developers risky, complex, and time-consuming [118]; such policy goes beyond the goals of ontologies.
3) The absence of a normalized and referenced method, leaving it to the individual modeler to find the best ways to represent things in a given domain [118].
4) The implementation of such ontologies requires a lot of manpower and time to manually integrate them [119].

To address some of these limitations our proposition was developed based on ontology design patterns. It will permit our work to be flexible, expandable, and re-usable [27], as well as to reduce the cost[14], improve the profitability, and manage the risks. Our proposal, thus, is the keystone of credit risk management platform.

4. 3.The Development of Credit Risk Scorecard Pattern

In recent years, ontologies have been increasingly used in industry and research fields for different purposes. The development of semantic domain ontology can help in reducing common problems and ambiguities linked to tag based systems [5]. Ontologies representing the domain knowledge have been used to guide the design of the application and to supply the system with semantic technologies [6]–[9]. The general ontology that models the credit risk management process and two specific ontologies have been proposed [120]. One of these specific models the process of credit allocation to clients, while the second displays necessary concepts for monitoring of a credit system. Ontologies are beneficial when they are used in decision support systems [10]. The latter work proposes an integrated ontological model for evaluating client applications which incorporates both: the default risk of investment and the development component of the investment. [121] proposes an approach for the conceptualization and the definition of business rules present in governance policies of the Brazilian financial system, more specifically those related to risk management. It proposes an ontology called Onto-Bacen that expresses the concepts (and their relationships) of this

[14]According to the 2015 McKinsey report, it is estimated that the implementation of BCBS 239 principles would cost approximately 230 million USD per G-SIB and 75 million USD in each D-SIB.

domain, and by using inference algorithms, which can verify the compliance of hypothetical financial institutions with those policies.

Decision making is one of the main research themes of systems' science, and DSSs were developed in many areas; e.g., management decision making, group decision making, etc. Authors in [122] developed an ontology based on decision support systems to assess the risk factors and provide appropriate treatment suggestions for diabetic patients. [123] presents an ontology-driven method for multi-criteria decision making that explicitly focuses on ensuring that the consequences of each choice are considered. In [124] the authors combine between mobile technology and fuzzy ontology and group decision-making algorithms, to facilitate the mobilization of knowledge, giving the user the possibility to get decision making support from the dynamic and massive data through their mobile services.

[125] attempts to present an ontological risk analysis in which the authors integrate three different perspectives on risk: (I) risk as a quantitative notion, (II) risk as a chain of events that impacts an agent's goals, (III) risk as the relationship of ascribing risk. Despite their effort, their proposal remains a superficial and simplified approach. Moreover, the implementation of this ontology in the analysis of risk in a specific domain, such as credit risk, remains a tedious task that may not succeed even though the life cycle of the development of the ontology is not specified and the ontology is not validated.

However, the above-mentioned solutions are not conclusive. Among their major drawbacks we can cite: requiring a lot of manpower and time to manually integrate them, designed and modeled for specific contexts with strong commitments and great details rendering their reusability and expandability difficult. To address some of these limitations we developed a flexible conceptual architecture based on ontology design patterns.

4.3.1.Requirements and Use Case

The use case that drives our modeling allows the credit analyst to calculate the number of points of each applicant (Table 4.1). It also provides them with a decision support tool (Table 4.2) which allows them to immediately reach an opinion on the credit allocation.

The modeling has to be as general as possible so that it becomes possible to add information from any sources (web, credit bureau data, financial ratios, social networks…etc.). The schema has to be robust, extendable, and easy to be maintained and updated. In order to achieve these objectives, the credit risk scorecard and decision support tool are modeled with ontology design patterns.

Table 4.1.*Example of credit scoring scorecard*

Variable	ClassVal0	Nbpoints
Age	>25 years <=25 years	0 8
Other_credits	No credit Other banks or institution	0 7
Accounts	No cheking account CA >= 200 euros CA [0-200 euros] CA < 0 euros	0 13 19 25
Credit_duration	<= 15 months 16-36 months > 36 months	0 13 18
Savings	No savings or >500 euros <500 euros	0 8
Guarantees	Guarantor No Guarantor	0 21
Credit_history	No credit at any time Credits without delay Creditswithnon-payments	0 6 13

Table 4.2. *Example of credit scoring decision support tool*

Nbpoints	Credit		
Frequency Percent Row Pct	Good	Bad	Total
Low risk	389 38.90 91.31	37 3.70 8.69	426 42.60
Medium risk	239 23.90 63.56	137 13.70 36.44	376 37.60
High risk	72 7.20 36.36	126 12.60 63.64	198 19.80
Total	700 70.00	300 30.00	1000 100.00

The ontology is built to address a set of requirements; in fact, it is evaluated against its corresponding requirements specification. These requirements can be defined through appropriate CQs, which define the scope of ontology. Some typical CQs are listed below:

1) Which variables take part in the credit risk scorecard?
2) Which credit risk scorecard does this variable participate in?
3) What is the value of a given category?
4) What is the score band of low, medium and high scores?
5) What does good and bad client mean?
6) How many counts of a client type are bad?
7) How many counts of a client type are good?
8) What is the risk of an applicant who is under 25 applying for credit for the first time at the institution with no other credit, no non-payments, with an account having slightly positive balance (but less than €200), with a small amount of savings (less than €500), and without a guarantor applying for credit for 36 months?
9) What is the risk of an applicant aged over 25 with credits in competing institutions, without non-payments, with an account having an average balance of more than €200, with more than €500 in savings, and without a guarantor applying credit for 12 months?

4.3.2.Data sources

Once we have a set of CQs as listed previously, we take a closer look at the data sources and the data structure. The used data set is credit risk scorecard (Table 4.1) and credit scoring decision support tool (Table 4.2) [126] which are the result of credit scoring modeling by logistic regression applied to ‘German credit data'[15]. The latter contains 1000 consumer credit files of which 700 ‘Good’ applicants (no non-payments) and its 300 ‘Bad’ applicants (non-payments) and 19 independent variables.

The credit risk scorecard contains the selected variables, their division into categories (ClassVal0) and the weight per category for each variable (nbpoints). The decision support tool is divided in three score bands (low risk, medium risk, high risk); each band contains the number and the percentage of good and bad applicants.

[15]https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data)

The credit risk scorecard will make it possible to respond to some CQs, e.g., questions (1) and (2). However, the credit risk scorecard alone will not allow the possibility of addressing CQs as in numbers (5), (6), (7) and (8) requiring the decision support tool, and (4) which would require more fine-grained information about the objectives of the project for which the credit risk scorecard is developed.

For additional information, possible sources are plentiful, mainly when the applicant's file needs close analysis and examination, e.g., credit bureau data, social network, demographic data. For the moment, attention will be given to model existing sources, credit risk scorecard and decision support tool. However, at the same time our modeling should be expandable as possible to include other data sources.

4.3.3.Credit Risk Scorecard Modeling

The system consists of two main ontologies credit risk scorecard and applicant ontologies. Since it is the same approach for two ontologies, we will focus on credit risk scorecard ontology. This ontology is the core of the system. It encapsulates all the information required to analyze the applicant information, examines the risks, computes risk score, and suggests credit decision, according to the scorecard guideline and applicant data.

4.3.3.1.Ontology modeling

We have a working group with a mix of participants: domain expert, participants specializing in database, and ontology engineer working on ODP-based modeling process. Reconciling the differing perceptions of domain expert to their topic, credit risk scorecard and decision support tool, - as data providers - and the CQs cited above leads to an expandable scope. Thus, requires data integration which is part of our use case. The modeling of our ontology will make the data publication simpler and graph structures intuitive, as well as making its reusability easier [127].

Based on the credit risk scorecard and decision support tool formats discussed earlier, and based on our CQs, we can identify several key notions which will be in need of modeling: agent, event, variables, categories, decision support tool, credit risk scorecard support tool, and credit risk scorecard.

Let us start with the notion of the Agent. In order to borrow from best practices and realize that "being an agent" is actually a role which an agent can take (e.g. Developer, Analyst, and Manager Etc.) for a certain period of time. We, therefore, reuse the common ontology design pattern for this purpose, which is depicted in Figure 4.1 in exactly the form in which it is used, e.g., in [128]. We opt for adapting this pattern for our specific case, leaving most of the things untouched, while avoiding overgeneralizations. The resulting pattern is depicted in Figure 4.2. Note that we have introduced three different agent roles: Analyst, Developer, and Manager. The yellow frames indicate that a more complex entity (a pattern in its own right) could stand in place of the frame. The latter could be modeled with a more fine-grained model, or alternatively, an existing ODP could be directly used instead.

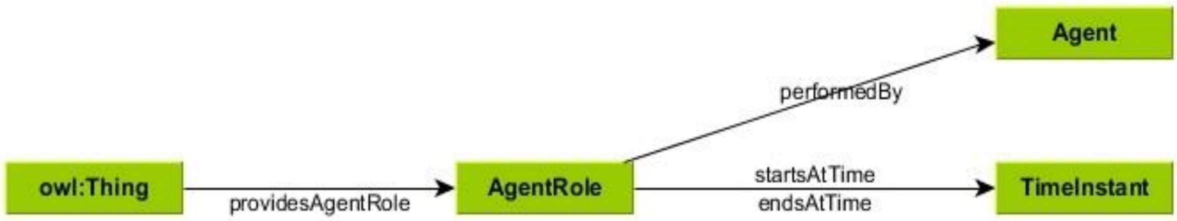

Figure 4.1. *The agent role Content ODP's UML graphical representation.*

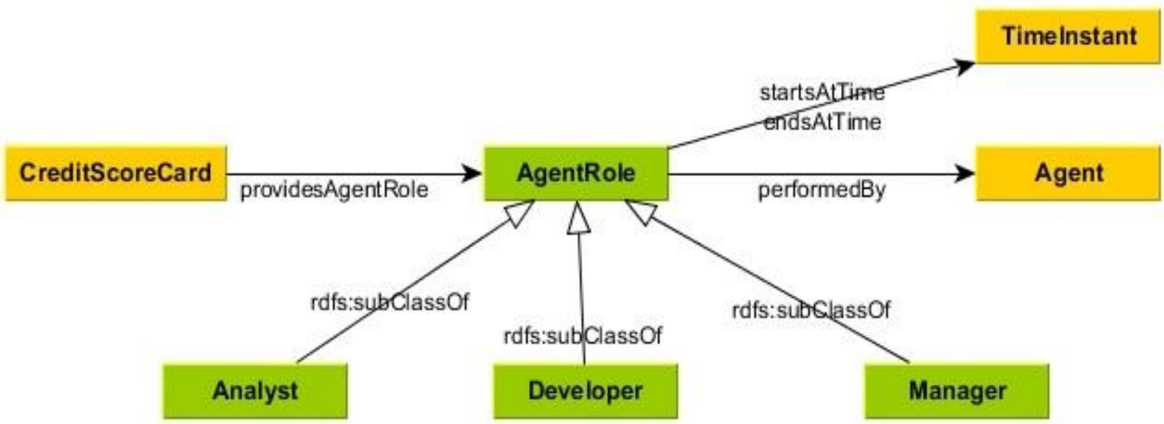

Figure 4.2. *Adapted agent role ODP for credit risk scorecard player roles UML graphical representation.*

Let us turn our attention to credit risk scorecard, from the viewpoint of the team's domain expert, the edition and validation of credit risk scorecard is an event, or alternatively as a step of credit risk scorecard development Figure 4.3. Therefore, we suggest to reuse a generic event pattern such as the one depicted in Figure 4.4, in exactly the form in which it is used, e.g. in [128].

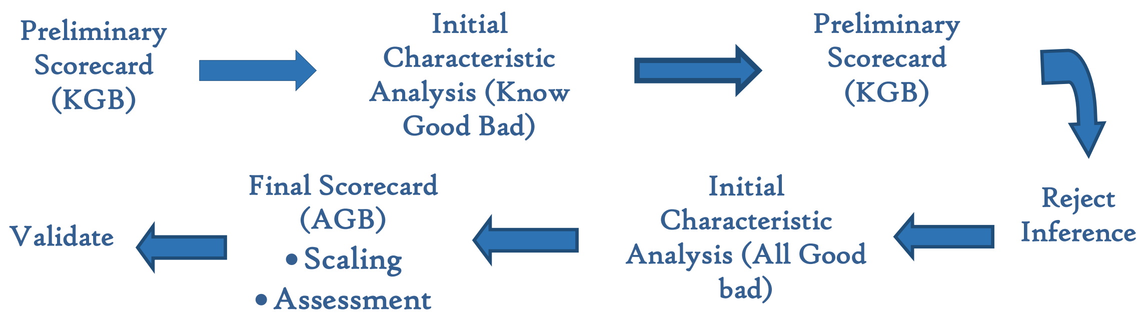

Figure 4.3. *Credit risk scorecard development steps.*

Figure 4.4. *The event Content ODP's UML graphical representation.*

As stated earlier, we adapt the pattern for our specific purpose, the result of which is depicted in Figure 4.5. Note that we linked directly the place to string containing the name of the place in which the credit risk scorecard was developed. The development of credit risk scorecard can be in-house or by external vendors. In contrast, if we had directly linked credit risk scorecard to the string containing the name of modelization algorithm, we would not have to provide an extension in the future. At the same time, we reuse the agent role pattern. However, the sub-event Figure 4.4 and temporal information **are not treated at this work**, yet we intend to provide the possibility to make this extension later without having to change anything already modeled.

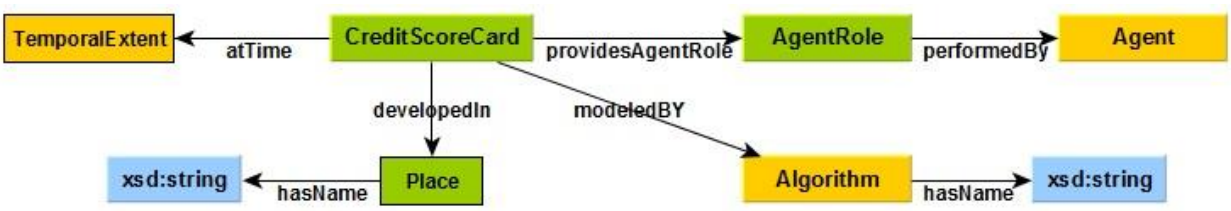

Figure 4.5. *The credit risk scorecard as event.*

The next step is dealing with variables; since we have settled for viewing the credit risk scorecard as event, the variables are naturally participants in the event. Therefore, we suggest reusing a generic participation pattern such as the one depicted in Figure 4.6 in exactly the form in which it is used[16]. Following the same approach, we adapt this pattern to fit our specific case, so we prefer to leave most of the things untouched. The resulting pattern is depicted in Figure 4.7.

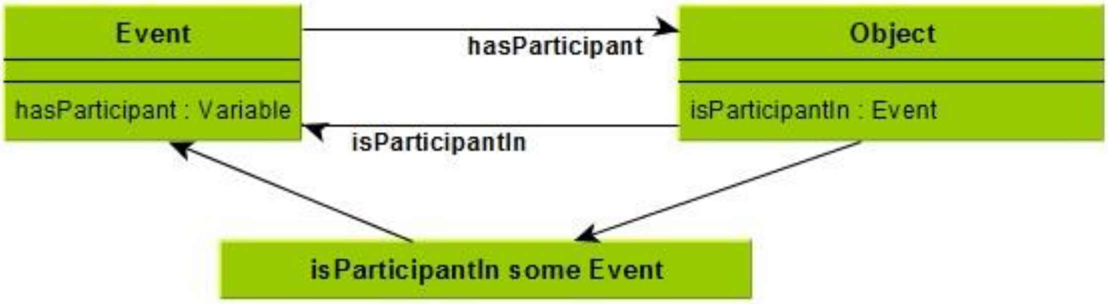

Figure 4.6. *The participation Content ODP's UML graphical representation.*

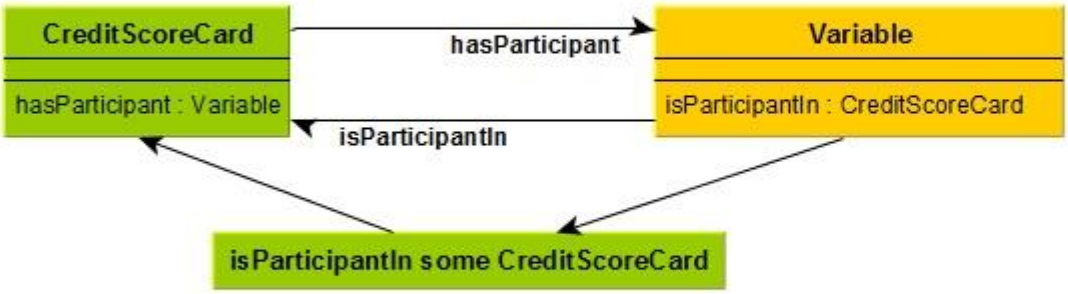

Figure 4.7. *The variables as participation.*

Considered from the viewpoint of our team's domain expert, the categories are classifications of variables. Therefore, we suggest reusing a generic classification pattern such as the one depicted in Figure 4.8 in exactly the form in which it is used[17]. We intend to adapt this pattern for our specific case. The resulting pattern is depicted in Figure 4.9.

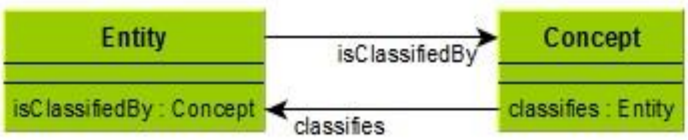

Figure 4.8. *The classification content ODP's UML graphic representation.*

[16] http://ontologydesignpatterns.org/wiki/Submissions:Participation

[17] http://ontologydesignpatterns.org/wiki/Submissions:Classification

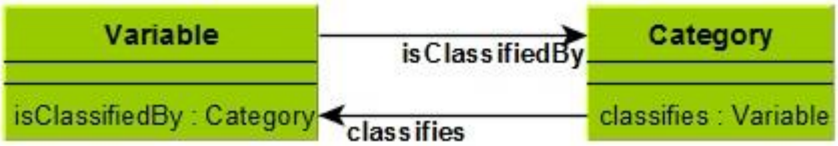

Figure 4.9. *The categories as classification.*

Moving to credit scoring decision support tool, the numbers of points are divided into bands. Therefore, the band risks (low, medium, high) are classifications of credit scoring decision support tool. Here, as in the previous case, we suggest to reuse a generic classification pattern depicted in Figure 4.8. The resulting pattern is depicted in Figure 4.10. It is worthy of note that we have introduced three different Band risks, two different Clients (Good and Bad) and the credit risk scorecard objectives. As mentioned above, and regarding future extension, we do not directly link client and credit risk scorecard objectives to strings containing the definition of client and credit risk scorecard objectives.

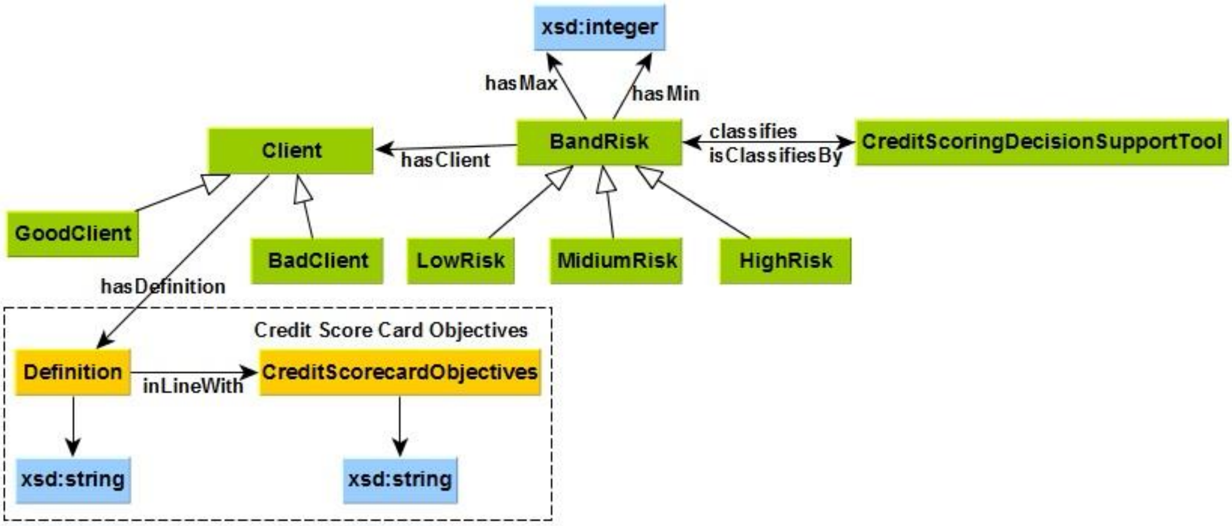

Figure 4.10. *The credit risk scoring decision support tool as a pattern.*

4.3.3.2.Relating Things Together

Finally, after the development of each piece of the whole credit risk scorecard, we intend to assemble all the pieces, producing the result depicted in Figure 4.11. Note that we directly linked credit risk scorecard to credit risk scorecard objectives and credit scoring decision support tool.

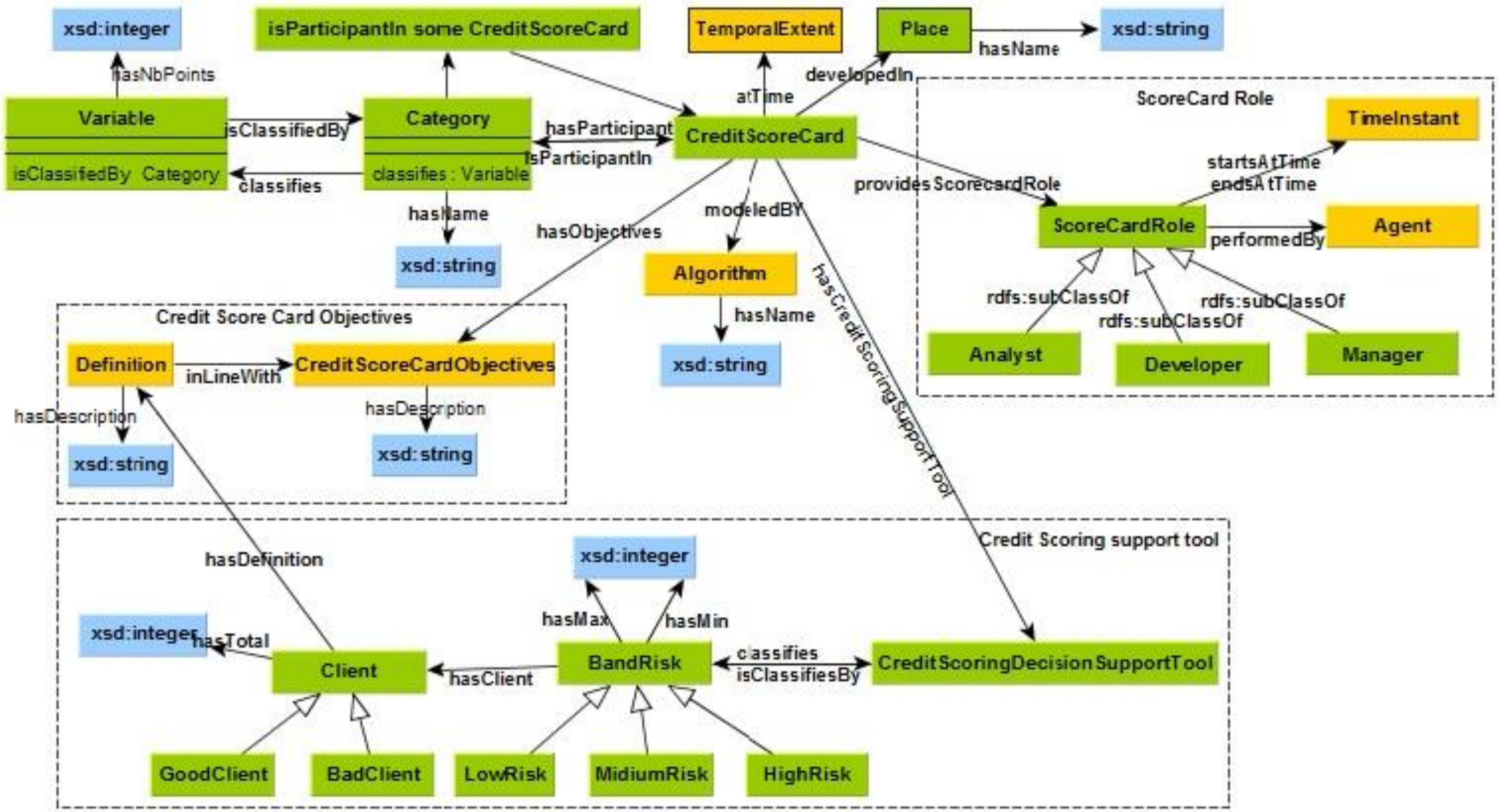

Figure 4.11. *The credit risk scorecard model.*

4. 4.Credit Risk Assessment

The system consists of two main ontologies, the credit risk scorecard ontology and the applicant profile ontology. They are implemented using formal language OWL, with the Protégé tool [129]. The proposed reasoning algorithm for the applicant analysis is illustrated in Figure 4.12.

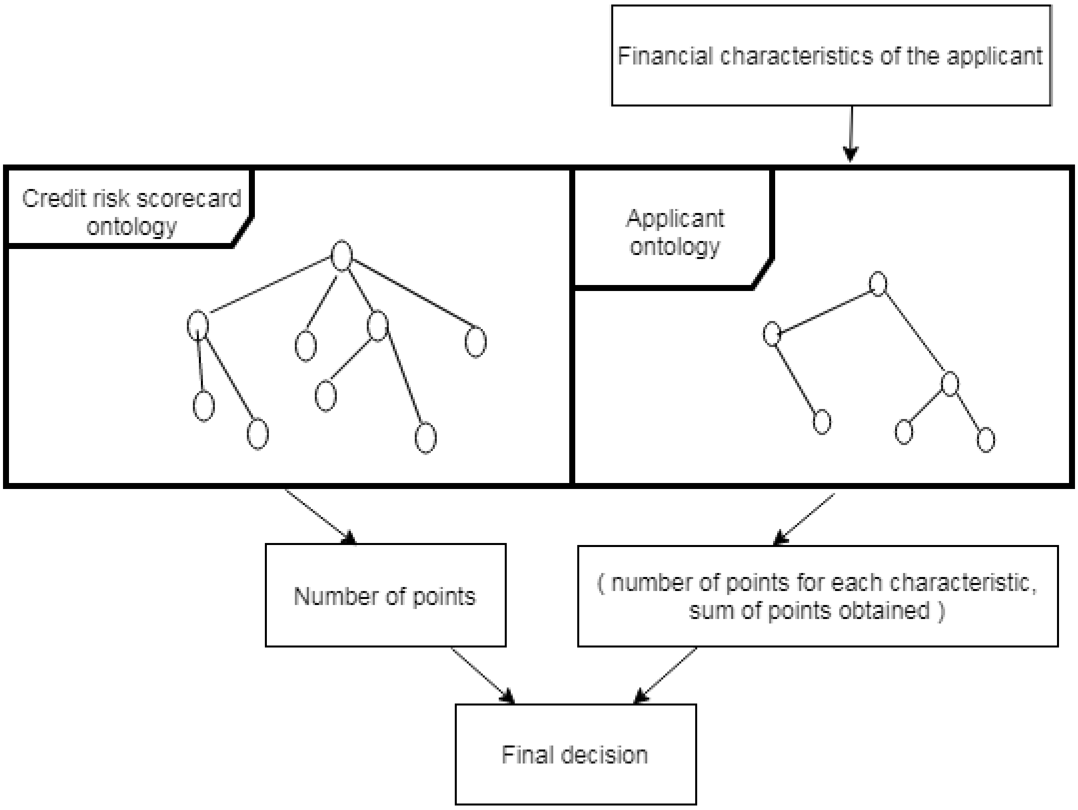

Figure 4.12. *Applicant analysis reasoning algorithm.*

4.4.1.Credit Risk Scorecard Ontology

Due to the lack of appropriate tools supporting ODPs in widely used IDEs and the absence of well-developed ontologies in credit risk field. We choose to develop a credit risk scorecard ontology from scratch based on ODPs.

4.4.1.1.Summary of Terminologies

We show the terminology used in credit risk scorecard ontology, which describes nineteen classes depicted in (Table 4.3), and nine of object properties in (Table 4.4).

Table 4.3. *Summary of classes used in credit risk scorecard ontology.*

Terminology name	Type	Definition
CreditScoreCardId	Class	A set of attributes that represents the ID of scorecard.
Definition	Class	A set of attributes that defines the client meaning. This must be in line with the credit scorecard objectives.
ScoreCardRole	Class	A set of attributes that defines the role of scorecard in the assessment process—is it a sole arbiter, or will it be used as a decision support tool?
Analyst	Class	A set of attributes that represents the analyst of the scorecard.
Manager	Class	A set of attributes that represents the manager of scorecard.
Developer	Class	A set of attributes that represents the developer of scorecard.
Algorithm	Class	A set of attributes that represents the algorithm used for constructing scorecard (linear regression, SVM, Decision tree etc.).
BandRisk	Class	A set of attributes that represents the band risk of applicant.
HighRisk	Class	A set of attributes that represents the type of client with high risk to not be granted the credit, meaning that the application is rejected.
MediumRisk	Class	A set of attributes that represents the type of client with medium risk to not be granted the credit, meaning that the file must be examined rather more closely.
LowRisk	Class	A set of attributes that represents the type of client with low risk to not be granted the credit.
CreditScoringDecisionSupportTool	Class	A set of attributes that allows the credit analyst to reach an opinion immediately on the customer that they are examining.
Client	Class	A set of attributes that represents the type of

		client.
GoodClient	Class	A set of attributes that constitutes the definition of a good client, meaning that the applicant has been/may be granted their credit.
BadClient	Class	A set of attributes that constitutes the definition of a bad client, meaning that the applicant for credit has defaulted in the past.
Place	Class	A set attributes that represents the development place of the scorecard, whether it is in-house or by external vendors, and provides reasoning.
Variable	Class	A set of attributes that represents variables of scorecard.
Category	Class	A set of attributes that represents categories of each for variable of scorecard.
CreditScoreCardObjectives	Class	A set of attributes that represents the organizational objectives behind the use of the scorecard.

4.4.1.2.Graphic Representation of Credit Risk Scorecard Pattern

Figure 4.13 represents the graphical diagram of the credit risk scorecard pattern. Additionally, (Table 4.4) represents objects properties of credit risk scorecard pattern.

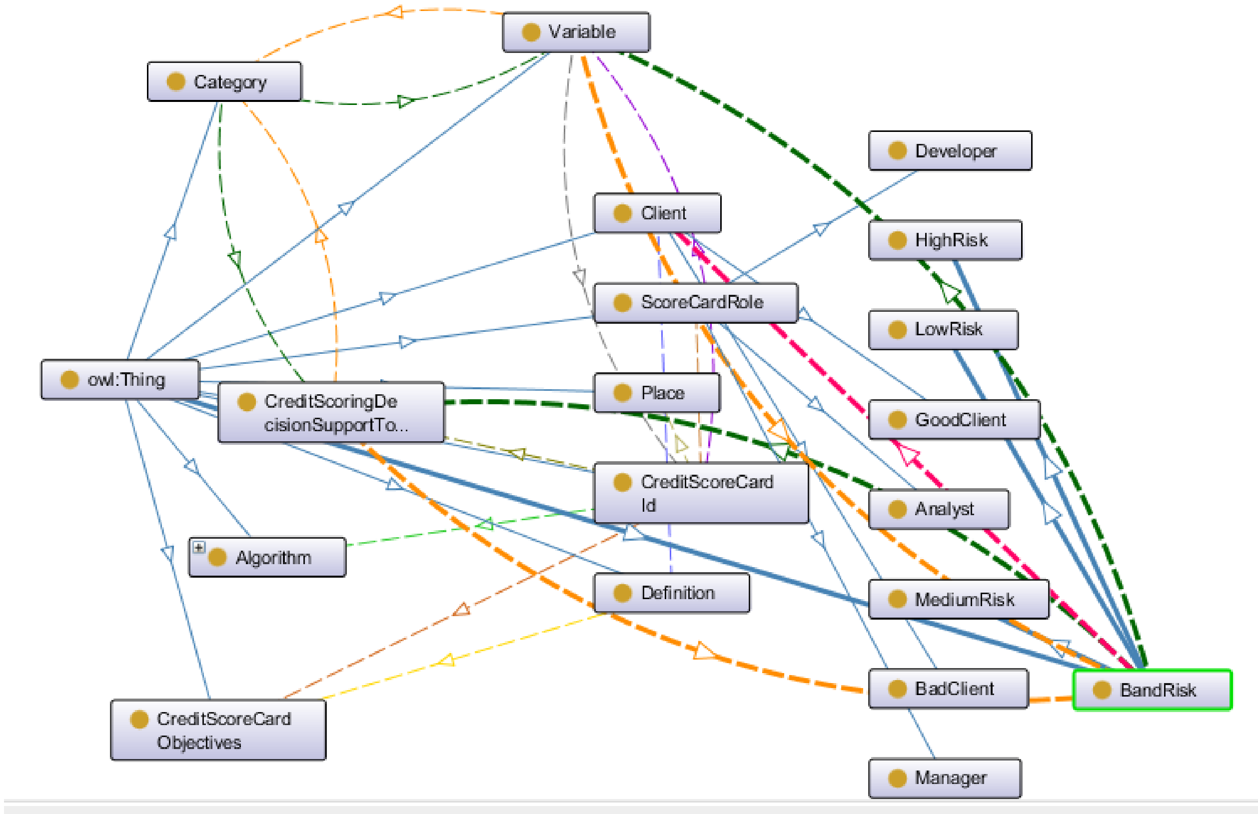

Figure 4.13. *Credit risk scorecard pattern graphical diagram.*

Table 4.4. *Summary of objects properties for credit risk scorecard content ontology design.*

Property name	Domain	Range	Inverse property
hasCreditScoringSupportTool	CreditScoreCardId	CreditScoringDecisionSupportTool	-
hasDefinition	Client	Definition	-
hasObjectives	CreditScoreCardID	CreditScoreCardObjectives	-
hasParticipant	CreditScoreCardID	Variable	isParticipant
developedIn	CreditScoreCardID	Place	-
inLineWith	Definition	CreditScoreCardObjectives	-
Classifies	Variable	Category	isClassifiedBy

Property name	Domain	Range	Inverse property
Categorises	CreditScoringDecisionSupporrTool	BandRisk	isCategorisedBy
providesAgentRole	CreditScoreCardID	ScoreCardRole	-

4.4.1.3.Class Hierarchy

Figure 4.14 represents the hierarchy diagram of classes of the credit risk scorecard pattern. It represents the core of the system and contains information about the credit risk scorecard model as in Figure 4.11.

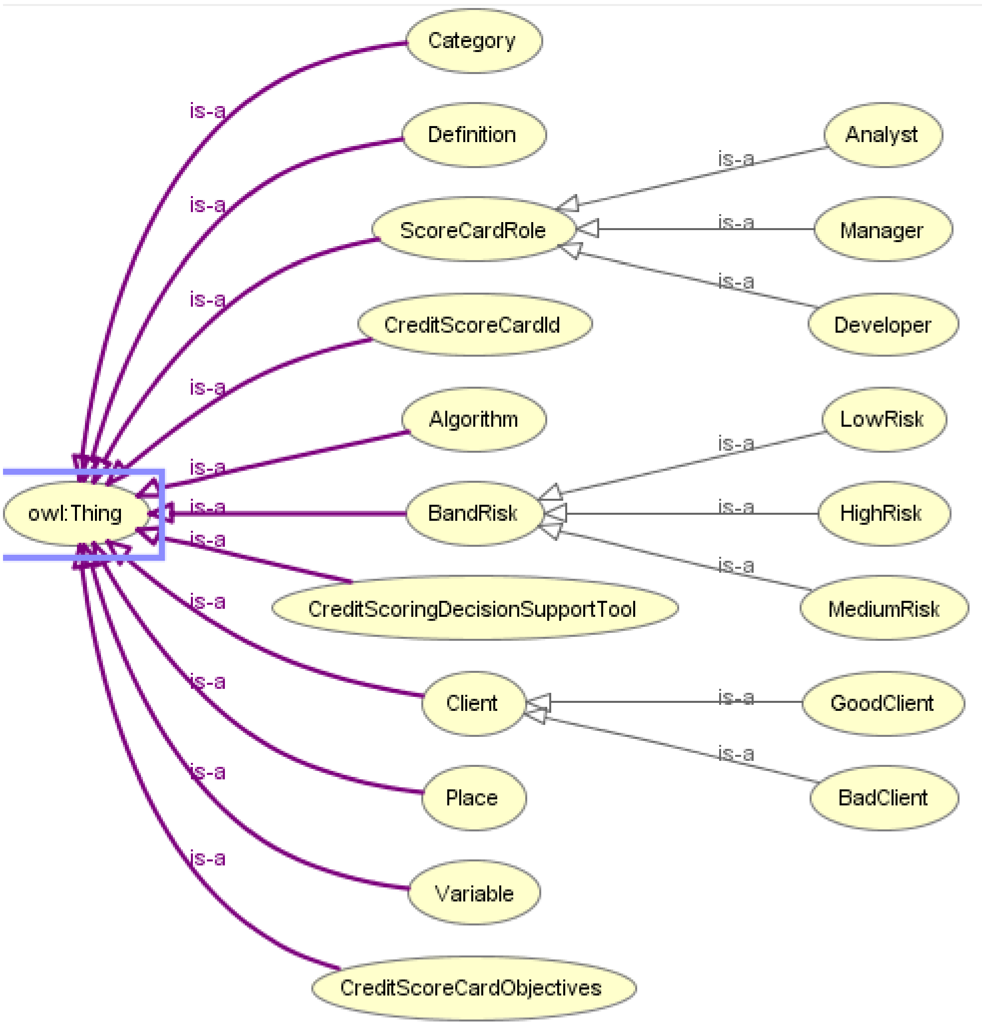

Figure 4.14. *Class hierarchies of credit risk scorecard ontology.*

4.4.2.Applicant Ontology

We choose to develop an applicant ontology from scratch based on ODPs. Applicant profile ontology, shown in Figure 4.15, is an OWL file that formalizes applicant details as entered by the credit risk scorecard analyst. This file is generated as soon as the applicant submits the loan.

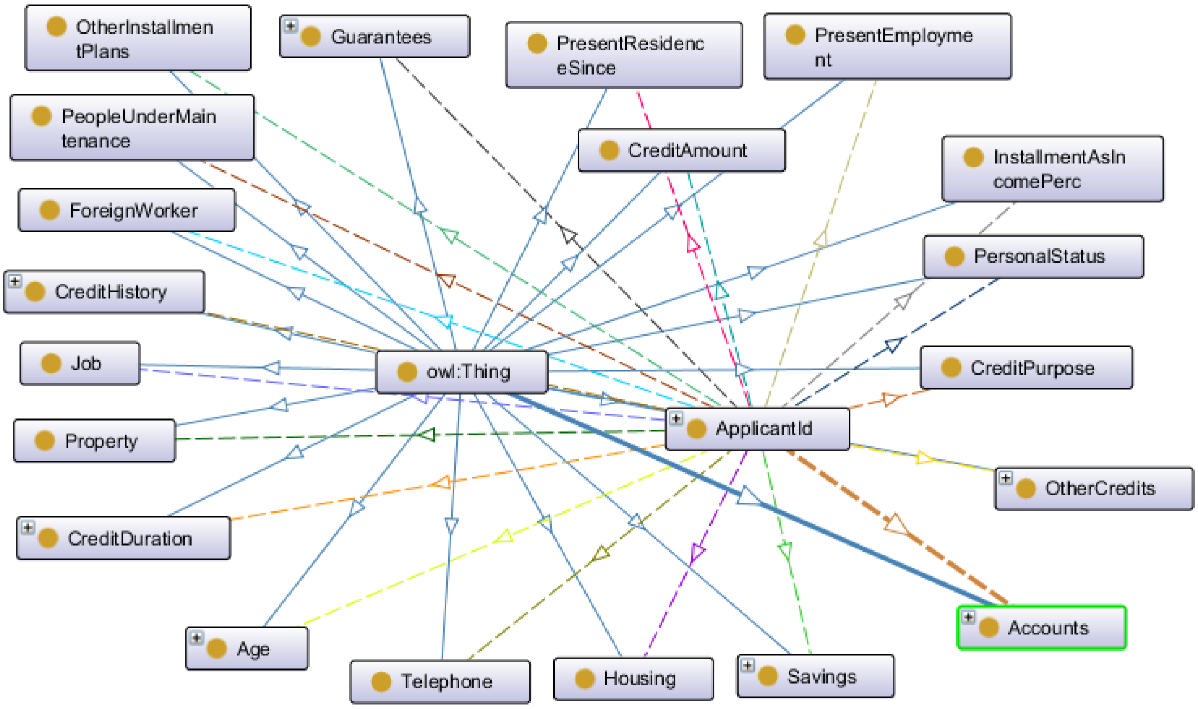

Figure 4.15. *Applicant pattern graphical diagram.*

4.4.2.1.Summary of Terminologies

The concept applicant profile has twenty-one classes described in (Table 4.5), and twenty-one objects properties described in (Table 4.6).

Table 4.5. *Summary of classes used in applicant ontology.*

Terminology name	Type	Definition
ApplicantId	Class	A set of attributes that represents the applicant.
Age	Class	A set of attributes that represents the age of applicant in years.

Terminology name	Type	Definition
Accounts	Class	A set of attributes that represents the status of account.
CreditAmount	Class	A set of attributes that represents the credit amount.
CreditDuration	Class	A set of attributes that represents the credit duration.
CreditHistory	Class	A set of attributes that represents the credit history of the applicant at this bank and other banks.
CreditPurpose	Class	A set of attributes that represents the purpose of credit.
ForeignWorker	Class	A set of attributes that represents whether or not the applicant is a foreign worker.
Guarantees	Class	A set of attributes that represents the guarantor.
Housing	Class	A set of attributes that represents the housing status of applicant.
InstallmentAsIncomePerc	Class	A set of attributes that represents installment rate in percentage of disposable income.
Job	Class	A set of attributes that represents the status and skills of applicant.
OtherCredits	Class	A set of attributes that represents the number of existing credits at this bank.
OtherInstallmentPlans	Class	A set of attributes that represents other installment plans.
PeopleUnderMaintenance	Class	A set of attributes that represents the number of people being liable to provide maintenance for.
PersonalStatus	Class	A set of attributes that represents personal status and sex.
PresentEmployment	Class	A set of attributes that represents the applicant’s current employment record.
PresentResidenceSince	Class	A set of attributes that represents the applicant’s current residence.
Property	Class	A set of attributes that represents properties other than savings.
Savings	Class	A set of attributes that represents savings account.
Telephone	Class	A set of attributes that represents whether or not the applicant has a telephone.

Table 4.6. *Summary of object properties for applicant content ontology design pattern.*

Property name	Domain	Range	Inverse property
has Accounts	ApplicantId	Accounts	-
has Age	ApplicantId	Age	-
hasCrediAmount	ApplicantId	CreditAmount	-
hasCreditDuration	ApplicantId	CreditDuration	-
hasCreditHistory	ApplicantId	CreditHistory	-
hasCreditPurpose	ApplicantId	CreditPurpose	-
hasForeignWorker	ApplicantId	ForeignWorker	-
hasGuarantees	ApplicantId	Guarantees	-
hasHousing	ApplicantId	Housing	-
hasInstallmentRateInPercentage	ApplicantId	InstallmentRateInPercentage	-
hasJob	ApplicantId	Job	-
hasNumberOfPeople	ApplicantId	PeopleUnderMaintenance	-
hasOtherCredit	ApplicantId	OtherCredit	-
hasOtherInstallmentPlans	ApplicantId	OtherInstallmentPlans	-
hasPersonalStatus	ApplicantId	PersonalStatus	-
hasPresesentEmployment	ApplicantId	PresentEmployment	-
hasPersonalStatus	ApplicantId	PersonalStatus	-
jasPresentResidenceSince	ApplicantId	PresentResidenceSince	-
hasProperty	ApplicantId	Property	-
hasSavings	ApplicantId	Savings	-
hasTelephone	ApplicantId	Telephone	-

4.4.2.2.Class Hierarchy

Figure 4.16 represents the hierarchy diagram of classes of the applicant pattern.

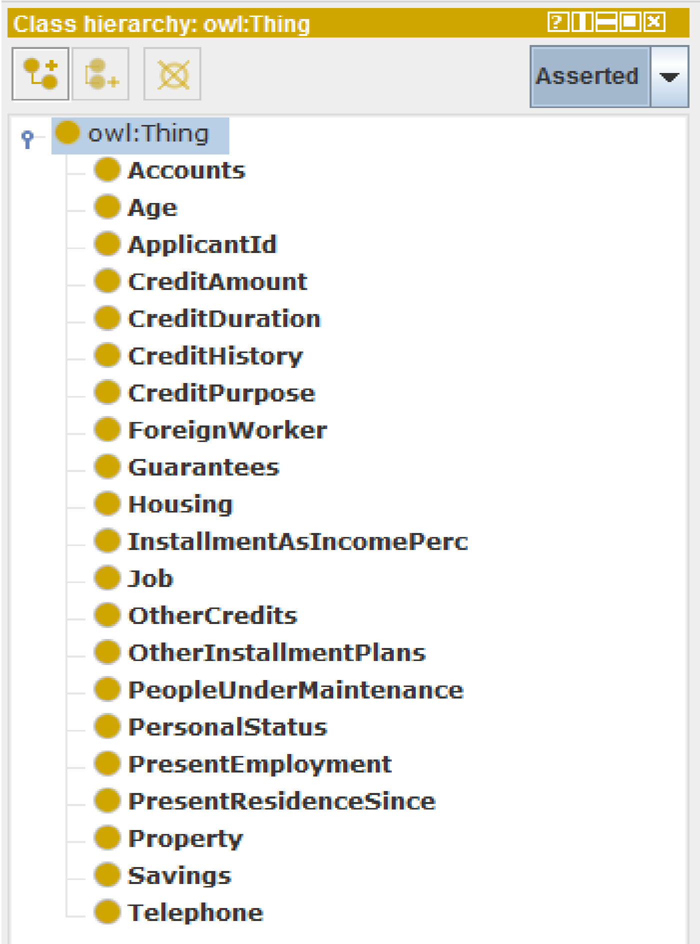

Figure 4.16. *Class hierarchies of applicant ontology.*

4. 5.Conclusion

In this chapter, we used the Krisnadhi & Hitzler methodology for modeling credit risk scorecard and applicant ontologies to improve the credit risk management. The modeling of credit risk scorecard ontology will directly link credit risk scorecard to credit risk scorecard objectives and credit risk scoring decision support tool. The ontology is then used to identify the corresponding tasks, subtasks, roles, actors, and resources of the affected business. Our proposal is expandable and aims at satisfying various needs. It provides modular, reusable, replaceable pieces, and it will make the data publication simpler and graph structures intuitive. Furthermore, it enables knowledge sharing both within and across organizations, which leads to better management of knowledge application. In the meantime, the ontologies that we have developed are available on the Web as OWL file, where the target community, particularly specific financial institutions (e.g. Credit bureau), may then re-use and expand them. To enrich our proposal in the next chapter, we will define ontologies classes (e.g. credit risk scorecard) via axiomatization. The axiomatic method constitutes a common framework for the discussion of scientific problems for people coming from different backgrounds and even for people working in different but strongly related branches of the same discipline (e.g. credit risk scorecard analyst, manager, and developer).

Chapter 5: Towards a Platform-Based Ontology Design Patterns and BCBS 239 for Credit Risk Management

Chapter 5 Towards a Platform-Based Ontology Design Patterns and BCBS 239 for Credit Risk Management

5. 1.Introduction

Within the past decades, significant advances have been made in the scorecard model for credit risk management. Practitioners and policy makers have invested a great deal in implementing and exploring a variety of new models individually. However, less progress has been made in coordinating and sharing information. As a consequence, for many banking systems, assertions or decisions that depend on data cannot be reliable. Motivated by specific challenges which received attention after the 2008 financial crisis, in this chapter we extend and validate the work developed in the previous chapter. That is, we fulfilled the second part of developing our system, which is collecting, reusing and sharing credit risk domain knowledge on the one hand, and, on the other, creating new knowledge application for improving the competitiveness of financial institutions. This will lead financial institutions to develop a credit risk management platform which conforms to the principles of BCBS 239.

We demonstrate the implementation of credit risk scorecard and applicant ontologies based on ODPs as a stepping stone towards a more complex and sophisticated framework for credit risk management and decision-making processes in financial institutions. The credit risk scorecard and applicant ontologies are extended, enriched using logical axioms, validated, distributed in OWL files, and checked in the test cases using SPARQL.

5. 2.Implementing the BCBS 239 Principles

In the aftermath of the 2008 financial crisis, it has become crystal clear that banks' information technology (IT) and data architectures were inadequate to support the broad management of financial risks [1]. That is, the crisis resulted primarily from the lack of ground-truth in many banking data assertions or the decisions that depend on the data, creating a confusion in terms used to manage risks. To improve the banks' ability to aggregate risk data[18], the BCBS issued an ensemble of principles (BCBS 239) which will lead to gains in efficiency, reduced probability of losses and enhanced strategic decision-making, and ultimately increased profitability, according to the BCBS 239.

[18] Risk data aggregation is defined by the BCBS 239 as "defining, gathering and processing risk data according to the bank's risk reporting requirements to enable the bank to measure its performance against its risk tolerance/appetite."

The BCBS 239 principles cover four main topics:

1) Overarching governance and infrastructure: this topic focuses on two principles, namely governance and data architecture and IT infrastructure requirements.
2) Risk data aggregation capabilities: accuracy and integrity, completeness, timeliness, and adaptability.
3) Risk reporting practices: accuracy, comprehensiveness, clarity and usefulness, frequency, and distribution.
4) Supervisory review, tools and cooperation: review, Remedial actions and supervisory measures, and home/host cooperation.

The paper [66] found that BCBS 239 is a unique opportunity to modernize IT platforms and enhance risk data management. It proposes the best practice solutions (Figure 5.1) for banks and other financial services to be able to implement them. In chapter 2, we provided some of the most crucial solutions for putting these best practices in place in order to address the issue of risk data management. The best approach available up to date for implementing these solutions is the use of ontologies in knowledge representation and reasoning (KR&R), knowledge sharing and reusing (KS&R), knowledge integration (KI), information retrieval (IR), decision support systems (DSSs), and user interfaces. What allows ontologies to be such a great approach for solving this problem is their use in each part of the information systems (Figure 5.1 below).

As presented in the Figure below, the end-user solutions are, in general, based on artificial intelligence in order to assess potential future scenarios, predict behavior in unprecedented ways, and extract meaningful information from the data. However, AI is not enough to reach the desired results since a good content theory of the domain is not implemented. In fact, as shown in [29], authors stated that once a good content theory is available, many different AI models might be used equally well to implement effective systems, all using essentially the same content. Thus, in order to deal with content theory and domain representation, a great interest was shown in ontologies.

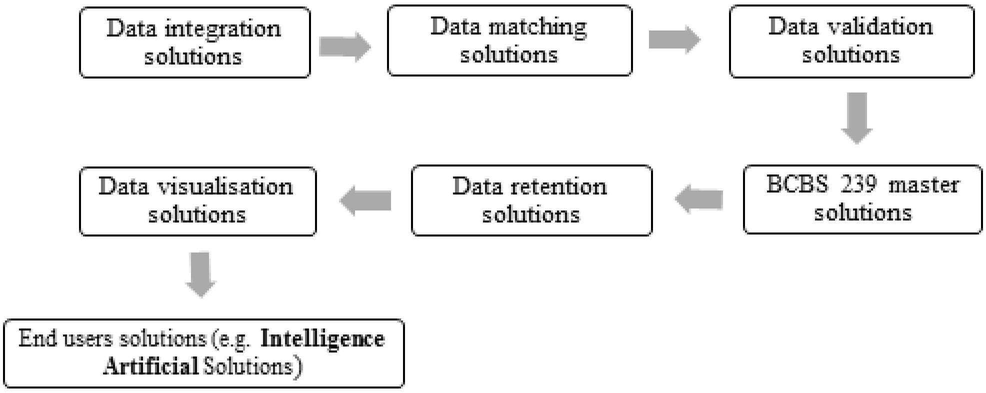

Figure 5.1. *Solutions for the implementation of BCBS 239.*

As defined in [16], an ontology is a formal, explicit specialization of shared conceptualization. Therefore, they provide potential terms for describing well the knowledge domain to benefit many different mechanisms such as rule systems, fuzzy logic, machine learning, deep learning, neural networks or natural language processing to be used and implemented using and sharing the same content.

Furthermore, building credit risk management ontologies on the basis of the open-world assumption (OWA), which means that our ontologies are built in a flexible manner to allow extension and correction, will not only enable us to implement the BCBS 239 principles but is also a great solution for incorporating the other measures and requirements proposed since the outbreak of the 2008 crisis such as Basel III and the Solvency II rules; revisions to the supervisory reporting frameworks of financial reporting (FINREP) and common reporting (COREP) as well as to the international financial reporting standards (IFRS) and to the Foreign Account Tax Compliance Act (FATCA). Taking into consideration upgraded risk data aggregation and risk reporting practices will allow banks to effectively comply with all these initiatives and requirements.

5. 3.Enrichment of Credit Risk Scorecard Pattern

In the last decade, the use of ontologies in the academia and industry fields has been exponentially expanded. They are used to manage and process the files/documents of various departments and sections of higher educational institutions in a paperless environment [130]. They are used in the healthcare field to provide navigable interactive healthcare guidance to assist ethnic minorities and aboriginal patients disadvantaged by languages, with limited use of the written text [131], and represent and update continuously information concerning gene products and their functions [132]. They are also used in renewable energy to help non-technical user to choose the domestic solar hot water system according to their needs, containing up-to-date information on its components and interrelationships, installation costs etc. [133]. They have also played a key role in the data management process from the knowledge representation [132] passing through the integration of heterogeneous data sources [50], [134], data cleaning [135] and the data publishing and access [136] ending with big data management [137] and access control for cloud data [138]. Additionally, ontologies have also been used in financial industry. FIBO is by far the sole and unique family of ontologies deployed in the field (this is talked in detail in the previous chapter). Nonetheless, the reusability of these ontologies is not without setbacks. Further discussion of these issues is detailed in chapter 3.

In this part we will develop expandable and re-usable credit risk scorecard and applicant ontologies, which will be the keystone of ontology-based credit risk management platform through the implementation of the BCBS 239 principles. This solution will permit to make good content theory and domain representation of the credit risk scorecard. It will also help financial institutions to develop a platform which conforms to the principles of BCBS 239 as well as provide user-ends solutions made available by AI mechanisms (e.g., machine learning and deep learning) which will be able to asses potential future scenarios, predict behavior in unprecedented ways, and extract meaningful information from the data. To realize this, we will extend and enrich the work developed in [139]. The latter has developed a credit risk scorecard and applicant ontologies based on ODPs. This work was the result of a group work by various experts: domain expert, specialists in data base addressed by modeling, and ontology engineer working on ODP-based modeling process. The reconciliation of the differing perceptions of domain experts to their topic, credit risk scorecard and decision support tool - as data providers - and the CQs have led to a flexible, expandable, and re-usable credit scorecard and applicant ontologies (see credit scorecard model in Figure 4.11).

Since the work on the credit risk scorecard and applicant ontologies is the same, we will only outline here the credit risk ontology in order to avoid repetition.

5.3.1.Axiomatization

The model depicted in Figure 4.11 is not the scorecard ontology, but it gives the representation of credit scorecard classes. Representation alone is not sufficient to sustain all the weight of the philosophy of science's building [140]. In [141]'s view, the best way to define this class of structure (e.g. credit risk scorecard) is via axiomatization.

The axiomatic method constitutes a common framework for the discussion of scientific problems for people coming from different backgrounds, and even for people working in different but strongly related branches of the same discipline (e.g. credit risk scorecard analyst, manager and developer). One of the multiples advantages of axiomatization is standardization. It can create standard terminology and standard methods of conceptual analysis in various branches of science[19]. A further advantage is that it has the value of the heuristic method. It can even be applied to disciplines in the process of formulation. Additionally, it plays crucial role in ontology modeling that aim to disambiguate the ontology.

For this purpose, we use axioms in description logics [142] to describe the pattern and translate them into OWL. The axioms in ontology do not provide complete information; thus, the description logics have been designed [142]. We know from [139] that the credit risk scorecard pattern consists of the following components:

1) The (re)use of the Agent Role pattern which reoccurs when modeling manager, analyst and developer of credit risk scorecard.
2) The (re)use of the Event pattern to model credit risk scorecard.
3) The (re)use of the Participation pattern to model credit risk scorecard variables.
4) The (re)use of the Classification Pattern to model categories of variables.
5) The credit scoring decision support tool pattern.

To reuse Agent role, Event, Participation and Classifications patterns in practice, there is only one way: importing the OWL serialization of four patterns into the OWL serialization of the credit risk scorecard. This means that all axioms and ontological commitments imposed by the patterns will be employed by credit scorecard pattern. Furthermore, the reuse of patterns may necessitate some adjustments and modifications following the use cases and modeling requirements. In the section we describe the most important patterns.

[19]One of the main uses of ontology in computer science is that of making explicit the intended meaning of a terminology. So, the effect towards finding a standard in the scientific practices that aim at explicitness present in the studies on ontology have both the goal of enhancing communication through the use of a formal ontology.

5.3.1.1.Agent Role Pattern

Axioms for the Agent Role pattern are depicted in Figure 5.2. The axioms (1), (2), (3), and (4) capture the domain, range and the scope domain, and range of each property in the pattern providesAgentRole, performedBy, startsAtTime and endsAtTime, (e.g.: The AgentRole and Agent entities provide the domain constraint and range respectively for the performedBy property). The axiom (5) asserts that every pair of classes amongst AgentRole, Agent and TimeInstant are pair wise disjoint. Hence, nothing can be simultaneous: an agent and an agent role, or an agent and time instant, or agent role and time instant.

$$\top \subseteq \forall \textbf{ providesAgentRole.AgentRole} \quad (1)$$
$$\exists \textbf{ performedBy.Agent} \subseteq \textbf{AgentRole} \quad (2)$$
$$\textbf{AgentRole} \subseteq \forall \textbf{ performedBy.Agent} \quad (3)$$
$$\textbf{AgentRole} \subseteq \forall \textbf{ starts.AtTime.TimeInstant} \sqcap \forall \textbf{ ends.AtTime.TimeInstant} \quad (4)$$
$$\textbf{DisjointClasses(AgentRole, Agent, TimeInstant)} \quad (5)$$

Figure 5.2. *Axiomatization for Agent role pattern.*

5.3.1.2.Event Pattern

Axioms for the Event pattern are depicted in Figure 5.3. Note that the event pattern imports the Agent Role pattern. Axiom (7) affirms that an event occurs at some temporal extent and some place. The axioms (8), (9) and (10) express the domain, range and scoped domain, and range for the properties: atTime and atPlace, while the axiom (10) asserts that every pair of classes amongst Event, TemporalExtent, Place, AgentRole and Agent are pair wise disjoint.

$$\textbf{import: (1), (2), (3), (4), (5)}$$
$$\textbf{Event} \subseteq \exists \textbf{ atPlace.Place} \sqcap \exists \textbf{ atTime.TemporalExtent} \quad (6)$$
$$\exists \textbf{ atTime.TemporalExtent} \sqcup \exists \textbf{ atPlace.Place} \quad (7)$$
$$\textbf{Event} \subseteq \forall \textbf{atTime.TemporalExtent} \sqcup \forall \textbf{atPlace.Place} \quad (8)$$
$$\textbf{DisjointClasse(Event, TemporalExtent, Place, AgentRole, Agent)} \quad (9)$$

Figure 5.3. *Axiomatization for Event pattern.*

5.3.1.3.Variable Pattern

Axioms for the Variable pattern are depicted in Figure 5.4. Note that the Variable pattern imports the participation pattern. The axioms (1), (2), (3), (4), (5), and (6) express the domain, range, and scoped domain for the properties: classifies and isClassifiedBy, while the axiom (7) asserts that classes Variable and Category are disjoint.

$$\top \subseteq \forall\ \mathbf{isClassifiedBy.Category} \tag{1}$$
$$\exists\ \mathbf{isClassifiedBy.Category} \subseteq\ \mathbf{Variable} \tag{2}$$
$$\mathbf{Category} \subseteq \forall\ \mathbf{isClassifiedBy.Variable} \tag{3}$$
$$\top \subseteq \forall\ \mathbf{classifies.Variable} \tag{4}$$
$$\exists\ \mathbf{classifies.Variable} \subseteq\ \mathbf{Category} \tag{5}$$
$$\mathbf{Variable} \subseteq \forall\ \mathbf{classifies.Category} \tag{6}$$
$$\mathbf{DisjointClasse(Category,\ Variable)} \tag{7}$$

Figure 5.4. ***Axiomatization for variable pattern.***

5.3.1.4.Implementation of Credit Risk Scorecard Axioms

As our team is comprises multi-domain, experts, we choose to use a visual tool [143] to implement credit risk scorecard axioms. This tools permit to generate most of axioms and add them automatically to the ontology (see Figure 5.5). Some axioms for credit risk scorecard pattern are imported from Agent Role, Event, Variable, Category, and Credit risk scoring decision support tool patterns.

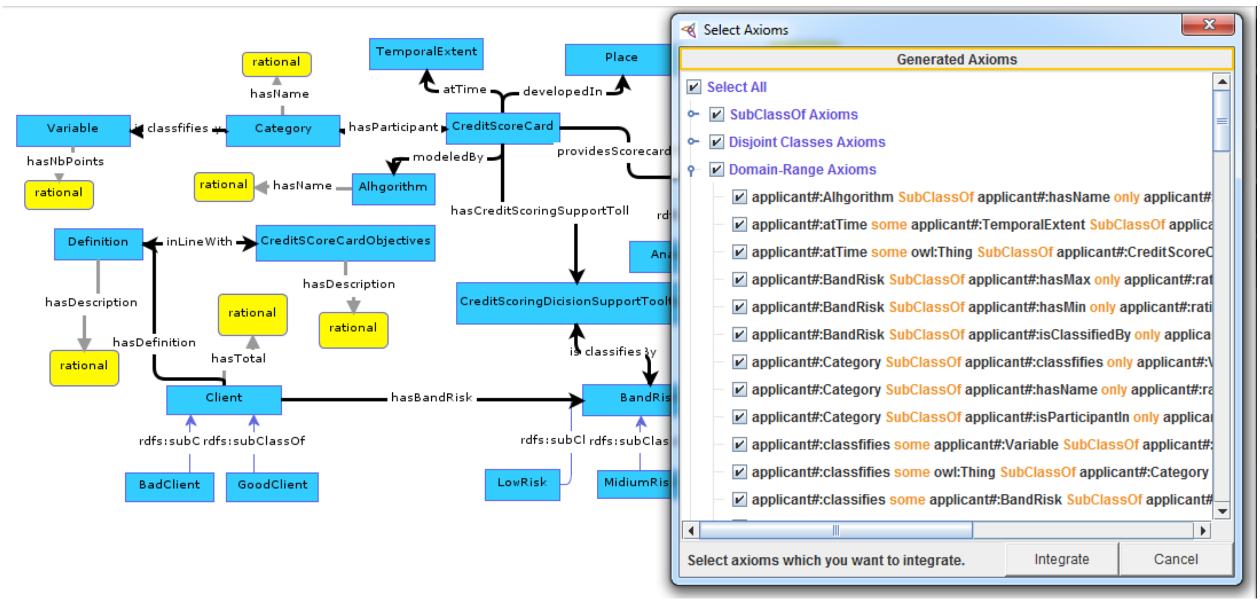

Figure 5.5. *The implementation of credit risk scorecard ontology axioms using OWLAx plugin [143].*

5.3.2.Validation of Credit Risk Scorecard Ontology

Ontology evaluation is a complex and tedious task. It includes the ontology diagnosis and repair [144]. In order to determine which ontologies would satisfy the requirements of a particular domain, the ontology has to be evaluated following an ensemble of criteria [145]: (i) accuracy: "a criterion that states if the definitions, descriptions of classes, properties, and individuals in an ontology are correct"; (ii) completeness: "measures if the domain of interest is appropriately covered in this ontology", (iii) conciseness: "the criteria that states if the ontology includes irrelevant elements with regards to the domain to be covered", (iv) adaptability "measures how far the ontology anticipates its uses. An ontology should offer the conceptual foundation for a range of anticipated tasks", (v) clarity "measures how effectively the ontology communicates the intended meaning of the defined terms. Definitions should be objective and independent of the context", (vi) computational efficiency "measures the ability of the used tools to work with the ontology, in particular the speed that reasoners need to fulfil the required tasks", and (vii) consistency "describes that the ontology does not include or allow for any contradictions". These criteria can be grouped in four categories:

1) Gold Standard-based approaches: they are based on comparing the developed ontology with a reference ontology that covers significantly a particular domain known as a gold ontology. They are efficient in evaluating the accuracy of ontology.
2) Corpus-based approaches: they are based on comparing the developed ontology with a reference corpus. They are efficient in evaluating the accuracy, completeness, and conciseness of ontology.
3) Task-based approaches: they are based on evaluating the performance of ontology according to a particular task for which the ontology is intended. They are efficient in evaluating the adaptability of an ontology.
4) Criteria based: they are efficient in evaluating the clarity of ontology using simple structure-based measures, or more complex measures.

In the recent years, enormous methods have been proposed for evaluating ontologies according to previous criteria. However, none of these methods can evaluate the ontology respecting all previous criteria [145]. Moreover, none of these methods are totally supported by tools [144]. Therefore, in this work we will use the methods and tools already tested. The validation of credit risk scorecard ontology is triply. Firstly, we make the diagnosis and repair of the credit risk scorecard ontology for quality validation. Secondly, we reason the logical consistency of the ontology. Thirdly, we make queries answering for usability validation of the ontology.

5.3.2.1.Quality validation

The validation of the quality of an ontology is an important part of ontology development. It is especially important when an ontology is expendable and re-usable.

We make the diagnosis and repair of credit risk scorecard ontology in order to check the technical quality of the ontology. However, manual diagnosis is always a tedious and time-consuming task. Therefore, we opt for using an automatic online evaluation framework (OntOlogy Pitfall Scanner, OOPS) [146] for OWL ontologies for many reasons: Online framework was accepted by the Semantic Web community, used by a high number of users worldwide, integrated by many systems, and it, further, enlarges the list of errors detected by most recent and available tools [144].

The credit risk scorecard ontology modeling does not contain any anomalies in which the quality of the ontology is checked through OOPS.

5.3.2.2.Consistency Validation

We mean by consistency ontology validation task the validation of the ontology through reasoner over three standard inference services [147]:

1) Consistency checking.
2) Classification.
3) Realization.

The credit risk scorecard ontology does not contain any contradictory facts in which logical consistency of the ontology is checked through HermiT [148] as one of the best reasoners [147] using its OWL API [149] version.

5.3.2.3.Usability Validation

We use query responses to validate the credit risk scorecard ontology usability. To achieve this objective we will use CQs as cited below at the assertion level by relying on SPARQL [150]. We have limited the validation of our ontologies to the selection of the two most fundamental CQs which are decisive in the credit loan decision:

- What is the risk of an applicant who is under 25 applying for credit for the first time at the institution with no other credit, no non-payments, with an account having slightly positive balance (but less than €200), with a small amount of savings (less than €500), and without a guarantor applying for credit for 36 months?

- What is the risk of an applicant aged over 25 with credits in competing institutions, without non-payments, with an account having an average balance of more than €200, with more than €500 in savings, and without a guarantor applying credit for 12 months?

A variety of tools are available as open source on which SPARQL can be executed and tested. For our case we choose Jana framework with ARQ [104]. Figure 5.6 and Figure 5.7 show the representation of the two applicants' data in turtle format and Figure 5.8 shows the representation of data credit scorecard model in turtle format.

```
# filename: applicant.ttl
@prefix ca: <http://ontocreditrisk.rf.gd/onto-creditrisk/ontology/applicant#> .
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

    ca:applicant ca:hasAge   "24"^^xsd:integer .
    ca:applicant ca:hasOtherCredit   "false"^^xsd:boolean .
    ca:applicant ca:hasAcounts "199"^^xsd:integer .
    ca:applicant ca:hasCreditDuration   "36"^^xsd:integer .
    ca:applicant ca:hasSavings   "400"^^xsd:integer .
    ca:applicant ca:hasGuarantees   "false"^^xsd:boolean .
    ca:applicant ca:hasCreditHistory   "No credit at any time"^^xsd:string .
    ca:applicant ca:applicant_id   "1"^^xsd:integer .
    ca:applicant ca:first_name   "jalil"^^xsd:string .
    ca:applicant ca:last_name   "elhassouni"^^xsd:string .
```

Figure 5.6. *The representation of applicant 1 in turtle format.*

```
# filename: applicant2.ttl
@prefix ca: <http://ontocreditrisk.rf.gd/onto-creditrisk/ontology/applicant#> .
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

    ca:applicant ca:hasAge   "26"^^xsd:integer .
    ca:applicant ca:hasOtherCredit   "Other banks or institution"^^xsd:string .
    ca:applicant ca:hasAcounts "201"^^xsd:integer .
    ca:applicant ca:hasCreditDuration   "12"^^xsd:integer .
    ca:applicant ca:hasSavings   "501"^^xsd:integer .
    ca:applicant ca:hasGuarantees   "No"^^xsd:string .
    ca:applicant ca:hasCreditHistory   "No credit at any time"^^xsd:string .
    ca:applicant ca:applicant_id   "2"^^xsd:integer .
    ca:applicant ca:first_name   "idriss"^^xsd:string .
    ca:applicant ca:last_name   "elhassouni"^^xsd:string .
```

Figure 5.7. *The representation of applicant 2 in turtle format.*

```
# filename: scorecard-data.ttl
PREFIX sc: <http://ontocreditrisk.rf.gd/onto-creditrisk/ontology/applicant#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
#variable
sc:v01 sc:name "Age" ;
       sc:classeId "1"^^xsd:integer ;
       sc:classeId "2"^^xsd:integer .
sc:v02 sc:name "Other_credits" ;
       sc:classeId "3"^^xsd:integer ;
       sc:classeId "4"^^xsd:integer .
sc:v03 sc:name "Accounts" ;
       sc:classeId "5"^^xsd:integer ;
       sc:classeId "6"^^xsd:integer ;
       sc:classeId "7"^^xsd:integer ;
       sc:classeId "8"^^xsd:integer .
sc:v04 sc:name "Credit_duration" ;
       sc:classeId "9"^^xsd:integer ;
       sc:classeId "10"^^xsd:integer ;
       sc:classeId "11"^^xsd:integer .
sc:v05 sc:name "Savings" ;
       sc:classeId "12"^^xsd:integer ;
       sc:classeId "13"^^xsd:integer ;
       sc:classeId "14"^^xsd:integer .
sc:v06 sc:name "Guarantees" ;
       sc:classeId "15"^^xsd:integer ;
       sc:classeId "16"^^xsd:integer .
sc:v07 sc:name "Credit_history" ;
       sc:classeId "17"^^xsd:integer ;
       sc:classeId "18"^^xsd:integer ;
       sc:classeId "19"^^xsd:integer .
```

Figure 5.8. *The representation of credit scorecard in turtle format – partial view.*

In the query below, Figure 5.9 calculates the total score of the two applicants and associates their credit risk:

```
# filename: score.rq
PREFIX ca: <http://ontocreditrisk.rf.gd/onto-creditrisk/ontology/applicant#>
PREFIX sc: <http://ontocreditrisk.rf.gd/onto-creditrisk/ontology/scorecarddata#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://example/>
PREFIX  dc:     <http://purl.org/dc/elements/1.1/>

SELECT ?id ?firstName ?lastName ?ageNote ?otherCreditNote
?accountsNote ?creditDurationNote ?savingsNote ?guaranteesNote ?creditHistoryNote
((?ageNote + ?otherCreditNote + ?accountsNote + ?creditDurationNote
+ ?savingsNote + ?guaranteesNote + ?creditHistoryNote) as ?totalNote)
?riskApp
FROM <applicant.ttl>
FROM <scorecard-data.ttl>
WHERE
{
    ca:applicant ca:hasAge ?age .
    ca:applicant ca:hasOtherCredit ?otherCredit .
    ca:applicant ca:hasAcounts ?accounts .
    ca:applicant ca:hasCreditDuration ?creditDuration .
    ca:applicant ca:hasSavings ?savings .
    ca:applicant ca:hasGuarantees ?guarantees .
    ca:applicant ca:hasCreditHistory ?creditHistory .
    ca:applicant ca:applicant_id ?id .
    ca:applicant ca:first_name ?firstName .
    ca:applicant ca:last_name ?lastName .

#Age
    sc:1    sc:min ?ageMin1 ;
            sc:max ?ageMax1 ;
            sc:nbPoint ?agePoint1 .
    sc:2    sc:min ?ageMin2 ;
            sc:max ?ageMax2 ;
```

Figure 5.9. *SPARQL query executed in Jana framework with ARQ – partial view.*

The following output is obtained after executing the code above:

ageNote	otherCreditNote	accountsNote	creditDurationNote	savingsNote	guaranteesNote	creditHistoryNote	totalNote	riskApp
8	0	19	13	8	21	0	69	"Medium risk"

Figure 5.10. *Answer query executed in Jana framework with ARQ over data scorecard and applicant 1.*

ageNote	otherCreditNote	accountsNote	creditDurationNote	savingsNote	guaranteesNote	creditHistoryNote	totalNote	riskApp
0	7	13	0	0	21	0	41	"Low risk"

Figure 5.11. *Answer query executed in Jana framework with ARQ over data scorecard and applicant 2.*

From the answer queries in Figure 5.10 and Figure 5.11 over credit risk scorecard and applicant's data, we infer that our ontologies are valid for utilization. Therefore, our ontologies are ready for publication and use by the community.

5. 4.Procedure to Submit Credit Risk Scorecard Ontology Pattern in ODPs Portal

In order to promote our ontology among the community users, we work on submitting our proposed pattern into the ontology design patterns portal[20]. Our aim also is to give a chance to our ontology to be reused, ameliorated, and expanded.

5. 5.Conclusion

This work identifies and formalizes classes and relations in credit risk scorecard and applicant ontologies. It demonstrates that these ontologies are suitable for describing the data contents of credit risk scorecard and applicants. Our work, however, is not limited to solely having an appropriate representation of the credit risk domain knowledge. It is also concerned with providing the right people with the proper knowledge that will assist them in making optimal decisions. This is done in such a way as to ensure that our shareable ontologies can be understood by humans and computers alike and at the same time remain open for knowledge accumulation and polymerization.

In this chapter, we have laid the foundation for ontology-based credit risk management platform in order to enhance credit risk management and decision-making process in financial institutions. We developed a rich, expandable, and re-usable credit risk scorecard and applicant ontologies which are based on ODPs to improve the decision-making process in credit loan. This will lead financial institutions to develop a credit risk management platform which conforms to the principles of BCBS 239. Both ontologies were validated and distributed in OWL files, and checked in the test cases using SPARQL. We extended the

[20] http://ontologydesignpatterns.org/wiki/Special:ImportProposal

ontologies which have been developed in chapter 4 and enriched them using logical axioms. They were triply validated; the authors:

4) Made diagnosis and repair of credit risk scorecard ontology in order to check the technical quality of the ontology.
5) Reasoned logical consistency of the ontology (consistency checking, classification, and realization).
6) Made queries answering for usability validation of the ontology.

To facilitate the re-use and the extension of these ontologies, we made sure they were distributed in OWL files.

Later on, our focus will be on ontology-based credit risk scorecard development. Thereafter, we can use Artificial Intelligence (machine learning and deep learning) to predict behavior of borrowers and new applicants in unprecedented ways and in real time in order to minimize the risks of financial institutions and enhance their competitivity in the financial market.

Conclusion and Future Work

Conclusion and Future Work

This section sums up the core elements of our work, first including the achievements-concerning the definition of ontology, the development of ontology engineering methodology, the modeling of scorecard, and the development of our ontologies, second the limitations of this research, and third the most important research prospects.

1. *Achievements*

The first accomplishment of this work is devoted to the definition of Ontology. We opted for a context-based definition that is crafted as the objectives demand, taking into consideration the problematic aspects of defining an ontology.

The second accomplishment we achieved in this dissertation is the choice of a proper methodology for developing our ontologies. We reviewed the most prominent ontology engineering methodologies, especially in a critical domain such as credit risk management, before our choice was set on the Krisnadhi & Hitzler methodology which is based on ontology design patterns.

The third accomplishment in this work is the conceptualization and modeling of the credit risk scorecard and applicant ontologies using the Krisnadhi & Hitzler methodology. Our ontologies provide modular, reusable, replaceable pieces, and will make the data publication simpler and graph structures intuitive. The ontologies are available on the Web as OWL file, where the target community, particularly specific financial institutions, may then re-use and expand them.

The fourth and final accomplishment is the development of a rich, expandable, and re-usable credit risk scorecard and applicant ontologies which are based on ODPs and in accordance with the principles of BCBS 239. Both ontologies were validated and distributed in OWL files, and checked in the test cases using SPARQL.

2. Limitations

Like any other promising dissertation, this research is not without limitations. The shortcomings of this work, at least the ones we are aware of, can be summarized in the following notes:

1) The main limitation of this work, at least in its theoretical and conceptual part, is that

there is no standard definition for the ontology. The ontology definition is often context-based and governed by the project objectives.

2) According to our review of the most prominent ontology engineering methodologies in chapter 3, the most mature and suitable approach for the development of ontology, especially in a critical domain such as credit risk management is Krisnadhi & Hitzler methodology. However, the main weaknesses of this methodology are: (i) it proposes the development of ontologies based only on OPDs, rather than on ontological and non-ontological resources, and (ii) as most other methodologies, it does not document well the process of ontology development, especially the extraction of terminologies.
3) Our single and unique data source was the credit risk scorecard and credit scoring decision support tool tables (see Table 4.1 and Table 4.2), which are the result of credit scoring modeling by logistic regression applied to ‘German credit data'. We were unable to reach other data sources, the data population of our ontology has not yet reached the threshold to justify ontology design model.
4) Our ontologies need to be adopted and used by the community to be tested and ameliorated. The degree of an ontology’s success and validation depends on its adoption by the target community.

3. *Perspectives*

All of the reviewed and proposed approaches in this thesis open new perspectives and future research directions. There are many directions in which this research can proceed, which can be categorized into the following areas:

1) Working out a definition of ontology that is neither too specific for a specific context nor too broad that it lacks applicable engineering process.
2) Working on a new ontology engineering methodology which combines both the philosophical and the engineering aspects of developing an ontology.
3) Automating the creation of scorecard-based ontology in different fields.
4) Building a platform-based Blockchain technology, ontology, and BCBS 239 for credit risk management.
5) Developing credit risk management upper-ontology using ISO standard (e.g. BFO-ISO).

References

[1] Basel Committee on Banking Supervision, *Principles for effective risk data aggregation and risk reporting*. Basel: Bank for Internat. Settlements, 2013.

[2] H. A. Abdou and J. Pointon, “Credit scoring, statistical techniques and evaluation criteria: a review of the literature,” *Intell. Syst. Account. Finance Manag.*, vol. 18, no. 2–3, pp. 59–88, 2011.

[3] B. DuCharme, *Learning SPARQL: querying and updating with SPARQL 1.1*. O’Reilly Media, Inc., 2013.

[4] M. Knight, “Taxonomy vs Ontology: Machine Learning Breakthroughs,” *DATAVERSITY*, Oct. 17, 2017. https://www.dataversity.net/taxonomy-vs-ontology-machine-learning-breakthroughs/ (accessed Oct. 02, 2020).

[5] M. Alruqimi and N. Aknin, “Enabling social WEB for IoT inducing ontologies from social tagging,” *Int. J. Inform. Commun. Technol.*, vol. 8, pp. 19–24, 2019.

[6] S. B. Kotsiantis, D. Kanellopoulos, V. Karioti, and V. Tampakas, “An ontology-based portal for credit risk analysis,” in *2009 2nd IEEE international conference on computer science and information technology*, 2009, pp. 165–169.

[7] H. Ahuja and R. Sivakumar, “Implementation of FOAF, AIISO and DOAP ontologies for creating an academic community network using semantic frameworks,” *Int. J. Electr. Comput. Eng.*, vol. 9, no. 5, p. 4302, 2019.

[8] R. Gunawan and K. Mustofa, “Finding Knowledge from Indonesian Traditional Medicine using Semantic Web Rule Language.,” *Int. J. Electr. Comput. Eng. 2088-8708*, vol. 7, no. 6, 2017.

[9] F. Sidi, I. Ishak, and M. A. Jabar, “MalayIK: An Ontological Approach to Knowledge Transformation in Malay Unstructured Documents.,” *Int. J. Electr. Comput. Eng. 2088-8708*, vol. 8, no. 1, 2018.

[10] S. Arsovski, B. Markoski, P. Pecev, L. Ratgeber, and N. Petrov, “An ontology driven credit risk scoring model,” in *2014 IEEE 15th International Symposium on Computational Intelligence and Informatics (CINTI)*, 2014, pp. 301–305.

[11] M. C. Suárez-Figueroa, A. Gómez-Pérez, and M. Fernández-López, “The NeOn Methodology framework: A scenario-based methodology for ontology development,” *Appl. Ontol.*, vol. 10, no. 2, pp. 107–145, 2015, doi: 10.3233/AO-150145.

[12] R. Arp, B. Smith, and A. D. Spear, *Building ontologies with basic formal ontology*. Mit Press, 2015.

[13] “Definitions of Ontology. First Part: from Wolff to Husserl.” https://www.ontology.co/ontology-definitions-one.htm (accessed Jun. 26, 2020).

[14] C. Feilmayr and W. Wöß, “An analysis of ontologies and their success factors for application to business,” *Data Knowl. Eng.*, vol. 101, pp. 1–23, Jan. 2016, doi: 10.1016/j.datak.2015.11.003.

[15] R. Neches *et al.*, “Enabling Technology for Knowledge Sharing,” *AI Mag.*, vol. 12, no. 3, Art. no. 3, Sep. 1991, doi: 10.1609/aimag.v12i3.902.

[16] T. R. Gruber, “A translation approach to portable ontology specifications,” *Knowl. Acquis.*, vol. 5, no. 2, pp. 199–220, Jun. 1993, doi: 10.1006/knac.1993.1008.

[17] W. N. Borst, “Construction of Engineering Ontologies for Knowledge Sharing and Reuse,” PhD Thesis, Center for telematica and information technlogy University of Tweently, 1997.

[18] R. Studer, V. R. Benjamins, and D. Fensel, “Knowledge engineering: Principles and methods,” *Data Knowl. Eng.*, vol. 25, no. 1, pp. 161–197, Mar. 1998, doi: 10.1016/S0169-023X(97)00056-6.

[19] S. Jean, G. Pierra, and Y. Ait-Ameur, “Domain Ontologies: A Database-Oriented Analysis,” in *Web Information Systems and Technologies*, Berlin, Heidelberg, 2007, pp. 238–254, doi: 10.1007/978-3-540-74063-6_19.

[20] N. Guarino, “Understanding, building and using ontologies,” *Int. J. Hum.-Comput. Stud.*, vol. 46, no. 2, pp. 293–310, Feb. 1997, doi: 10.1006/ijhc.1996.0091.

[21] H. Paulheim and F. Probst, “Ontology-Enhanced User Interfaces: A Survey,” *Int. J. Semantic Web Inf. Syst.*, vol. 6, no. 2, pp. 36–59, Apr. 2010, doi: 10.4018/jswis.2010040103.

[22] C. Roussey, F. Pinet, M. A. Kang, and O. Corcho, “An Introduction to Ontologies and Ontology Engineering,” in *Ontologies in Urban Development Projects*, G. Falquet, C. Métral, J. Teller, and C. Tweed, Eds. London: Springer, 2011, pp. 9–38.

[23] G. van Heijst, A. Th. Schreiber, and B. J. Wielinga, “Using explicit ontologies in KBS development,” *Int. J. Hum.-Comput. Stud.*, vol. 46, no. 2, pp. 183–292, Feb. 1997, doi: 10.1006/ijhc.1996.0090.

[24] R. Mizoguchi, Y. Tijerino, and M. Ikeda, “Task analysis interview based on task ontology,” *Expert Syst. Appl.*, vol. 9, no. 1, pp. 15–25, Jan. 1995, doi: 10.1016/0957-4174(94)00044-V.

[25] I. Niles and A. Pease, “Towards a Standard Upper Ontology,” in *Formal Ontology in Information Systems: Collected Papers from the Second International Conference*, 2001, pp. 2–9, doi: 10.1145/505168.505170.

[26] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider, “Sweetening Ontologies with DOLCE,” in *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*, Berlin, Heidelberg, 2002, pp. 166–181, doi: 10.1007/3-540-45810-7_18.

[27] P. Grenon, B. Smith, and L. Goldberg, “Biodynamic Ontology: Applying BFO in the Biomedical Domain,” *Stud. Health Technol. Inform.*, vol. 102, pp. 20–38, 2004, doi: 10.3233/978-1-60750-945-5-20.

[28] N. B. Cocchiarella, “Can an AI System Think? Functionalism and the Nature of Mentality,” presented at the Incontri 2019 su Scienza e Pensiero: La Complessità, Benevento, Mar. 2019.

[29] B. Chandrasekaran, J. Josephson, and V. R. Benjamins, “What Are Ontologies, and Why Do We Need Them?,” *Intell. Syst. Their Appl. IEEE*, vol. 14, pp. 20–26, 1999, doi: 10.1109/5254.747902.

[30] B. Chandrasekaran, “AI, Knowledge, and the Quest for Smart Systems,” *IEEE Expert Intell. Syst. Their Appl.*, vol. 9, no. 6, pp. 2–5, Dec. 1994, doi: 10.1109/64.363254.

[31] E. Bolisani and C. Bratianu, “The Elusive Definition of Knowledge,” in *Emergent Knowledge Strategies: Strategic Thinking in Knowledge Management*, E. Bolisani and

C. Bratianu, Eds. Cham: Springer International Publishing, 2018, pp. 1–22.

[32] R. Brachman and H. Levesque, *Knowledge Representation and Reasoning*. Amsterdam ; Boston: Morgan Kaufmann Publishers In, 2004.

[33] R. J. Brachman, *A Structural Paradigm for Representing Knowledge*. Harvard University, 1977.

[34] M. R. Quillian, *Semantic memory*. Air Force Cambridge Research Laboratories, Office of Aerospace Research, United States Air Force, 1966.

[35] J. F. S. Jr, *Conceptual Structures: Information Processing in Mind and Machine*. Reading, Mass: Longman Higher Education, 1983.

[36] M. Minsky, "A Framework for Representing Knowledge," Jun. 1974, Accessed: Jun. 27, 2020. [Online]. Available: https://dspace.mit.edu/handle/1721.1/6089.

[37] X. Li, S. Zhang, R. Huang, B. Huang, C. Xu, and Y. Zhang, "A survey of knowledge representation methods and applications in machining process planning," *Int. J. Adv. Manuf. Technol.*, vol. 98, no. 9, pp. 3041–3059, Oct. 2018, doi: 10.1007/s00170-018-2433-8.

[38] O. Marino Drews, "Classification reasoning in an object oriented multiple viewpoints representation," Theses, Université Joseph-Fourier - Grenoble I, 1993.

[39] W. Gödert, "An Ontology-based Model for Indexing and Retrieval," *J. Assoc. Inf. Sci. Technol.*, vol. 65, 2013, doi: 10.1002/asi.23420.

[40] M. Á. Rodríguez-García, R. Valencia-García, F. García-Sánchez, and J. J. Samper-Zapater, "Ontology-based annotation and retrieval of services in the cloud," *Knowl.-Based Syst.*, vol. 56, pp. 15–25, Jan. 2014, doi: 10.1016/j.knosys.2013.10.006.

[41] N. H. Shah, C. Jonquet, A. P. Chiang, A. J. Butte, R. Chen, and M. A. Musen, "Ontology-driven indexing of public datasets for translational bioinformatics," *BMC Bioinformatics*, vol. 10 Suppl 2, p. S1, Feb. 2009, doi: 10.1186/1471-2105-10-S2-S1.

[42] R. Stevens, C. A. Goble, and S. Bechhofer, "Ontology-based knowledge representation for bioinformatics," *Brief. Bioinform.*, vol. 1, no. 4, pp. 398–414, Nov. 2000, doi: 10.1093/bib/1.4.398.

[43] P. Ribino, A. Oliveri, G. L. Re, and S. Gaglio, "A Knowledge Management System Based on Ontologies," in *2009 International Conference on New Trends in Information and Service Science*, Jun. 2009, pp. 1025–1033, doi: 10.1109/NISS.2009.105.

[44] H. Lamharhar, D. Chiadmi, and L. Benhlima, "Ontology-based knowledge representation for e-government domain," in *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*, Brussels, Belgium, Dec. 2015, pp. 1–10, doi: 10.1145/2837185.2837203.

[45] W. Abramowicz, E. Lewańska, M. Kaczmarek, and S. Monika, "Ontology Structure, Reasoning Approach and Querying Mechanism in a Semantic-Enabled Efficient and Scalable Retrieval of Experts," *Int. J. Adv. Softw.*, pp. 510–520, 2011.

[46] J. Davies, A. Duke, and A. Stonkus, "OntoShare: Using Ontologies for Knowledge Sharing," in *Proceedings of the WWW2002 International Workshop on the Semantic Web, Hawaii, May 7, 2002*, 2002, vol. 55, [Online]. Available: http://ceur-ws.org/Vol-55/davies.pdf.

[47] R. Costa, C. Lima, J. Sarraipa, and R. Jardim-Gonçalves, "Facilitating knowledge sharing and reuse in building and construction domain: an ontology-based approach," *J. Intell. Manuf.*, vol. 27, no. 1, pp. 263–282, Feb. 2016, doi: 10.1007/s10845-013-0856-5.

[48] X. Li, Z. Wu, M. Goh, and S. Qiu, "Ontological knowledge integration and sharing for collaborative product development," *Int. J. Comput. Integr. Manuf.*, vol. 31, no. 3, pp. 275–288, Mar. 2018, doi: 10.1080/0951192X.2017.1407876.

[49] P. Ziegler and K. R. Dittrich, "Data Integration — Problems, Approaches, and Perspectives," in *Conceptual Modelling in Information Systems Engineering*, J. Krogstie, A. L. Opdahl, and S. Brinkkemper, Eds. Berlin, Heidelberg: Springer, 2007, pp. 39–58.

[50] I. F. Cruz and H. Xiao, "The role of ontologies in data integration," *Eng. Intell. Syst. Electr. Eng. Commun.*, vol. 13, no. 4, p. 245, 2005.

[51] H. Wache *et al.*, "Ontology-Based Integration of Information-A Survey of Existing Approaches.," 2001.

[52] M. Alizadeh, M. H. Shahrezaei, and F. Tahernezhad-Javazm, "Ontology based information integration: a survey," *ArXiv Prepr. ArXiv190913762*, 2019.

[53] B. Smith *et al.*, "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration," *Nat. Biotechnol.*, vol. 25, no. 11, pp. 1251–1255, Nov. 2007, doi: 10.1038/nbt1346.

[54] H. Ning and T. Wang, "Constructing a Dataspace Based on Metadata and Ontology for Complicated Scientific Data Management," in *2007 2nd International Conference on Pervasive Computing and Applications*, Jul. 2007, pp. 512–514a, doi: 10.1109/ICPCA.2007.4365497.

[55] M. del M. R. García, J. García-Nieto, and J. F. Aldana-Montes, "An ontology-based data integration approach for web analytics in e-commerce," *Expert Syst. Appl.*, vol. 63, pp. 20–34, Nov. 2016, doi: 10.1016/j.eswa.2016.06.034.

[56] J. Chen, D. Dosyn, V. Lytvyn, and A. Sachenko, "Smart Data Integration by Goal Driven Ontology Learning," in *Advances in Big Data*, Cham, 2017, pp. 283–292, doi: 10.1007/978-3-319-47898-2_29.

[57] G. De Giacomo, D. Lembo, M. Lenzerini, A. Poggi, and R. Rosati, "Using Ontologies for Semantic Data Integration," in *A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years*, S. Flesca, S. Greco, E. Masciari, and D. Saccà, Eds. Cham: Springer International Publishing, 2018, pp. 187–202.

[58] D. Calvanese, P. Liuzzo, A. Mosca, J. Remesal, M. Rezk, and G. Rull, "Ontology-based data integration in EPNet: Production and distribution of food during the Roman Empire," *Eng. Appl. Artif. Intell.*, vol. 51, pp. 212–229, May 2016, doi: 10.1016/j.engappai.2016.01.005.

[59] C. D. Manning, H. Schütze, and P. Raghavan, *Introduction to information retrieval*. Cambridge : Cambridge University Press, 2008.

[60] W. U. Dan and W. Hui-lin, "Role of ontology in information retrieval," *J. Electron. Sci. Technol.*, vol. 4, no. 2, pp. 148–154, 2006.

[61] V. Vijayarajan, M. Dinakaran, P. Tejaswin, and M. Lohani, "A generic framework for ontology-based information retrieval and image retrieval in web data," *Hum.-Centric Comput. Inf. Sci.*, vol. 6, no. 1, p. 18, Nov. 2016, doi: 10.1186/s13673-016-0074-1.

[62] K. Munir and M. Sheraz Anjum, "The use of ontologies for effective knowledge modelling and information retrieval," *Appl. Comput. Inform.*, vol. 14, no. 2, pp. 116–126, Jul. 2018, doi: 10.1016/j.aci.2017.07.003.

[63] V. Lytvyn, V. Vysotska, D. Dosyn, O. Lozynska, and O. Oborska, "Methods of Building Intelligent Decision Support Systems Based on Adaptive Ontology," in *2018 IEEE Second International Conference on Data Stream Mining Processing (DSMP)*, Aug. 2018, pp. 145–150, doi: 10.1109/DSMP.2018.8478500.

[64] M. Soui, S. Diab, A. Ouni, A. Essayeh, and M. Abed, "An Ontology-Based Approach for User Interface Adaptation," in *Advances in Intelligent Systems and Computing: Selected Papers from the International Conference on Computer Science and Information Technologies, CSIT 2016, September 6-10 Lviv, Ukraine*, N. Shakhovska, Ed. Cham: Springer International Publishing, 2017, pp. 199–215.

[65] M. Kultsova, A. Potseluico, and A. Dvoryankin, "Ontology Based Personalization of Mobile Interfaces for People with Special Needs," in *Creativity in Intelligent Technologies and Data Science*, Cham, 2019, pp. 422–433, doi: 10.1007/978-3-030-29750-3_33.

[66] L. Prorokowski and H. Prorokowski, "Solutions for risk data compliance under BCBS 239," *J. Invest. Compliance*, 2015.

[67] J. Orgeldinger, "The Implementation of Basel Committee BCBS 239: Short analysis of the new rules for Data Management," *J. Cent. Bank. Theory Pract.*, vol. 7, no. 3, pp. 57–72, 2018.

[68] E. F. Kendall, D. L. McGuinness, Y. Ding, and P. Groth, *Ontology Engineering*. Miejsce nieznane: Morgan & Claypool Publishers, 2019.

[69] M. Keet, *An introduction to ontology engineering*, vol. 1. College Publications, 2018.

[70] E. F. Aminu, I. O. Oyefolahan, M. B. Abdullahi, and M. T. Salaudeen, "A Review on Ontology Development Methodologies for Developing Ontological Knowledge Representation Systems for various Domains.," *Int. J. Inf. Eng. Electron. Bus.*, vol. 12, no. 2, 2020.

[71] R. Subhashini and J. Akilandeswari, "A survey on ontology construction methodologies," *Int. J. Enterp. Comput. Bus. Syst.*, vol. 1, no. 1, pp. 60–72, 2011.

[72] M. C. Suarez-Figueroa *et al.*, "NeOn methodology for building contextualized ontology networks," *NeOn Deliv. D*, vol. 5, p. 150, 2008.

[73] R. Iqbal, M. A. A. Murad, A. Mustapha, and N. M. Sharef, "An analysis of ontology engineering methodologies: A literature review," *Res. J. Appl. Sci. Eng. Technol.*, vol. 6, no. 16, pp. 2993–3000, 2013.

[74] M. Grüninger and M. S. Fox, "Methodology for the design and evaluation of ontologies," 1995.

[75] O. Corcho, M. Fernández-López, and A. Gómez-Pérez, "Methodologies, tools and languages for building ontologies. Where is their meeting point?," *Data Knowl. Eng.*, vol. 46, no. 1, pp. 41–64, 2003.

[76] M. Fernández-López, A. Gómez-Pérez, and N. Juristo, "Methontology: from ontological art towards ontological engineering," 1997.

[77] A. S. Abdelghany, N. R. Darwish, and H. A. Hefni, “An Agile Methodology for Ontology Development,” *Int. J. Intell. Eng. Syst.*, vol. 12, no. 2, pp. 170–181, 2019.

[78] S. Peroni, “A simplified agile methodology for ontology development,” in *OWL: Experiences and Directions–Reasoner Evaluation*, Springer, 2016, pp. 55–69.

[79] J. Axelsson, E. Papatheocharous, J. Nyfjord, and M. Törngren, “Notes on agile and safety-critical development,” *ACM SIGSOFT Softw. Eng. Notes*, vol. 41, no. 2, pp. 23–26, 2016.

[80] P. Hitzler, “Modeling with ontology design patterns: Chess games as a worked example,” *Ontol. Eng. Ontol. Des. Patterns Found. Appl.*, vol. 25, p. 3, 2016.

[81] P. Hitzler and A. Krisnadhi, “A tutorial on modular ontology modeling with ontology design patterns: The cooking recipes ontology,” *ArXiv Prepr. ArXiv180808433*, 2018.

[82] C. Shimizu, Q. Hirt, and P. Hitzler, “Modl: A modular ontology design library,” *ArXiv Prepr. ArXiv190405405*, 2019.

[83] N. F. Noy and D. L. McGuinness, *Ontology development 101: A guide to creating your first ontology*. Stanford knowledge systems laboratory technical report KSL-01-05 and …, 2001.

[84] I. Jacobson, M. J. Christerson, and P. Jonsson, *P. & vergaard, G.(1992). Object-Oriented Software Engineering, A Use Case Driven Approach*. Addison-Wesley, Wokingham.

[85] I. Jacobson, I. Spence, and K. Bittner, *USE-CASE 2.0. The Guide to Succeeding with Use Cases, Ivar Jacobson International sa*. 2011.

[86] M. Uschold and M. Gruninger, “Ontologies: Principles, methods and applications,” *Tech. Rep.-Univ. Edinb. Artif. INTELLIGENCE Appl. Inst. AIAI TR*, 1996.

[87] E. Blomqvist and K. Sandkuhl, “Patterns in Ontology Engineering: Classification of Ontology Patterns.,” in *ICEIS (3)*, 2005, pp. 413–416.

[88] A. Gangemi, “Ontology design patterns for semantic web content,” in *International semantic web conference*, 2005, pp. 262–276.

[89] E. Blomqvist, P. Hitzler, K. Janowicz, A. Krisnadhi, T. Narock, and M. Solanki, “Considerations regarding Ontology Design Patterns.,” *Semantic Web*, vol. 7, no. 1, pp. 1–7, 2016.

[90] A. Gangemi and V. Presutti, “Ontology design patterns,” in *Handbook on ontologies*, Springer, 2009, pp. 221–243.

[91] K. Hammar, “Ontology design patterns in WebProtégé,” 2015.

[92] “Welcome to the NeOn Project.” http://neon-project.org/nw/Welcome_to_the_NeOn_Project.html (accessed Sep. 14, 2020).

[93] K. Hammar, “Ontology design patterns: Adoption challenges and solutions,” in *Second International Workshop on Semantic Web Enterprise Adoption and Best Practice in conjunction with the 11th European Semantic Web Conference (ESWC 2014), Anissaras, Greece, May 26, 2014*, 2014, pp. 23–32.

[94] “OPTypes - Odp,” *Ontology Design Patterns. org (ODP) - Odp*. D (accessed Sep. 14, 2020).

[95] M. keet, “A requirements catalogue for ontology languages,” *Keet blog*, Apr. 10, 2020. https://keet.wordpress.com/2020/04/10/a-draft-requirements-catalogue-for-ontology-languages/ (accessed Sep. 02, 2020).

[96] X. Su and L. Ilebrekke, “A comparative study of ontology languages and tools,” in *International Conference on Advanced Information Systems Engineering*, 2002, pp. 761–765.

[97] D. Kalibatiene and O. Vasilecas, “Survey on ontology languages,” in *International Conference on Business Informatics Research*, 2011, pp. 124–141.

[98] G. Klyne and J. J. Carroll, “Resource Description Framework (RDF): Concepts and Abstract Syntax.” https://www.w3.org/TR/rdf-concepts/ (accessed Aug. 30, 2020).

[99] S. Staab and R. Studer, *Handbook on ontologies*. Springer Science & Business Media, 2010.

[100] R. T. Herrera, “Towards RDF Normalization. (Normalisation des ressources RDF).,” PhD Thesis, University of Pau and Pays de l’Adour, France, 2016.

[101] M. Uschold, Y. Ding, and P. Groth, *Demystifying OWL for the Enterprise*, vol. 8. Morgan & Claypool Publishers, 2018.

[102] M. K. Smith, C. Welty, and D. L. McGuinness, “OWL Web Ontology Language Guide,” *OWL Web Ontology Language Guide*, Feb. 2004. https://www.w3.org/TR/2004/REC-owl-guide-20040210/#OwlVarieties (accessed Sep. 04, 2020).

[103] E. Prud’hommeaux and A. Seaborne, “SPARQL Query Language for RDF.” https://www.w3.org/TR/rdf-sparql-query/ (accessed Sep. 06, 2020).

[104] The Apache Software Foundation, “Apache Jena - ARQ - A SPARQL Processor for Jena.” https://jena.apache.org/documentation/query/ (accessed Sep. 08, 2020).

[105] M. Casu, G. Cicala, and A. Tacchella, “Ontology-based data access: An application to intermodal logistics,” *Inf. Syst. Front.*, vol. 15, no. 5, pp. 849–871, Nov. 2013, doi: 10.1007/s10796-012-9395-4.

[106] M. R. Khondoker and P. Mueller, *Comparing Ontology Development Tools Based on an Online Survey*. 2010.

[107] “protégé.” https://protege.stanford.edu/products.php (accessed Sep. 13, 2020).

[108] R. J. Feldman, “Small business loans, small banks and big change in technology called credit scoring,” *The Region*, pp. 19–25, 1997.

[109] N. Siddiqi, *Intelligent credit scoring: Building and implementing better credit risk scorecards*. John Wiley & Sons, 2017.

[110] N. Kritzinger and G. W. Van Vuuren, “An optimised credit scorecard to enhance cut-off score determination,” *South Afr. J. Econ. Manag. Sci.*, vol. 21, no. 1, pp. 1–15, 2018.

[111] G. Dong, K. K. Lai, and J. Yen, “Credit scorecard based on logistic regression with random coefficients,” *Procedia Comput. Sci.*, vol. 1, no. 1, pp. 2463–2468, 2010.

[112] A. Peussa, “Credit risk scorecard estimation by logistic regression,” University of Helsinki, Faculty of Science, Department of Mathematics and Statistics, Master’sThesis, 2016.

[113] R. Anderson, *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*. Oxford: OUP Oxford, 2007.

[114] "FIBO," *EDM Council*. https://spec.edmcouncil.org/fibo/ (accessed Sep. 25, 2020).

[115] D. E. Wisnosky and the FIBO Leadership Team, "FIBO™ Build, Test, Deploy and Maintain Methodology," EDM Council, Guide, 2018. [Online]. Available: https://spec.edmcouncil.org/fibo/doc/FIBO_BTDM.pdf.

[116] "Gene Ontology Resource," *Gene Ontology Resource*. http://geneontology.org/ (accessed Sep. 25, 2020).

[117] "TOVE Project | Enterprise Integration Laboratory – EIL." http://eil.utoronto.ca/projects/tove-project/ (accessed Sep. 25, 2020).

[118] M. G. Bennett, "Providing Conceptual Disambiguation for Terms in Reusable Ontologies: A Case Study from FIBO.," 2018.

[119] J. Zaino, "Banking on FIBO: Financial Institutions Turn to Semantic Standard - DATAVERSITY," *DATAVERSITY - Data Education for Business and IT Professionals*, 2016. https://www.dataversity.net/banking-fibo-financial-institutions-turn-standard-value-compliance/ (accessed Sep. 25, 2020).

[120] M. Haziti, A. Qadi, M. Bazzi, and J. Elhassouni, "Applying ontologies to data integration systems for bank credit risk management," *J. Data Min. Digit. Humanit.*, 2018.

[121] F. Polizel, S. J. Casare, and J. S. Sichman, "OntoBacen: A Modular Ontology for Risk Management in the Brazilian Financial System.," *JOWO IJCAI*, vol. 1517, 2015.

[122] P. C. Sherimon and R. Krishnan, "OntoDiabetic: an ontology-based clinical decision support system for diabetic patients," *Arab. J. Sci. Eng.*, vol. 41, no. 3, pp. 1145–1160, 2016.

[123] G. Mansingh and L. Rao, "Enhancing the Decision Making Process: An Ontology-based Approach.," in *CONF-IRM*, 2014, p. 9.

[124] J. A. Morente-Molinera, R. Wikström, E. Herrera-Viedma, and C. Carlsson, "A linguistic mobile decision support system based on fuzzy ontology to facilitate knowledge mobilization," *Decis. Support Syst.*, vol. 81, pp. 66–75, 2016.

[125] T. P. Sales, F. Baião, G. Guizzardi, J. P. A. Almeida, N. Guarino, and J. Mylopoulos, "The common ontology of value and risk," in *International Conference on Conceptual Modeling*, 2018, pp. 121–135.

[126] S. Tufféry, *Data mining and statistics for decision making*. John Wiley & Sons, 2011.

[127] P. Hitzler, K. Janowicz, and A. A. Krisnadhi, "Ontology Modeling with Domain Experts: The GeoVocamp Experience," in *1st International Diversity++ Workshopco-located with the 14th International Semantic Web Conference (ISWC 2015)*, Pennsylvania, USA, 2015, vol. 1501, pp. 31–36.

[128] A. Krisnadhi *et al.*, "The GeoLink modular oceanography ontology," in *International Semantic Web Conference*, 2015, pp. 301–309.

[129] H. Knublauch, R. W. Fergerson, N. F. Noy, and M. A. Musen, "The Protégé OWL plugin: An open development environment for semantic web applications," in *International semantic web conference*, 2004, pp. 229–243.

[130] S. K. Dwivedi and A. Kumar, "Development of University Ontology for aSPOCMS," *J. Emerg. Technol. Web Intell.*, vol. 5, no. 3, pp. 213–221, Aug. 2013, doi: 10.4304/jetwi.5.3.213-221.

[131] D. Forbes, P. Wongthongtham, J. Singh, and S. Thompson, "Ontology Supported Assistive Communications in Healthcare," *Commun. Assoc. Inf. Syst.*, vol. 34, no. 1, Jan. 2014, doi: 10.17705/1CAIS.03414.

[132] Gene Ontology Consortium, "The gene ontology resource: 20 years and still going strong," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D330–D338, 2019.

[133] E. Kontopoulos, G. Martinopoulos, D. Lazarou, and N. Bassiliades, "An ontology-based decision support tool for optimizing domestic solar hot water system selection," *J. Clean. Prod.*, vol. 112, pp. 4636–4646, 2016.

[134] P. Monnin, C. Jonquet, J. Legrand, A. Napoli, and A. Coulet, "PGxO: A very lite ontology to reconcile pharmacogenomic knowledge units," 2017.

[135] S. Brüggemann, "Rule Mining for Automatic Ontology Based Data Cleaning," in *Progress in WWW Research and Development*, Berlin, Heidelberg, 2008, pp. 522–527, doi: 10.1007/978-3-540-78849-2_52.

[136] M. Buranarach *et al.*, "OAM: An Ontology Application Management Framework for Simplifying Ontology-Based Semantic Web Application Development," *Int. J. Softw. Eng. Knowl. Eng.*, vol. 26, no. 01, pp. 115–145, Feb. 2016, doi: 10.1142/S0218194016500066.

[137] B. Eine, M. Jurisch, and W. Quint, "Ontology-Based Big Data Management," *Systems*, vol. 5, no. 3, Art. no. 3, Sep. 2017, doi: 10.3390/systems5030045.

[138] A. Michael, R. Kothandaraman, and K. Kaliyan, "Providing Ontology-Based Access Control for Cloud Data by Exploiting Subsumption Property among Domains of Access Control," 2019, doi: 10.22266/IJIES2019.0630.27.

[139] J. Elhassouni, A. El Qadi, M. Bazzi, and M. El Haziti, "Modeling with ontologies design patterns: credit scorecard as a case study," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 17, no. 1, Art. no. 1, Jan. 2020, doi: 10.11591/ijeecs.v17.i1. pp. 429-439.

[140] R. Ferrario, "Who Cares about Axiomatization? Representation, Invariance, and Formal Ontologies *," 2006.

[141] P. Suppes, "Heuristics and the Axiomatic Method," in *Models and Methods in the Philosophy of Science: Selected Essays*, P. Suppes, Ed. Dordrecht: Springer Netherlands, 1993, pp. 55–65.

[142] M. Krötzsch, F. Simancik, and I. Horrocks, "A description logic primer," *ArXiv Prepr. ArXiv12014089*, 2012, [Online]. Available: http://arxiv.org/abs/1201.4089.

[143] M. Sarker, A. A. Krisnadhi, and P. Hitzler, "OWLAx: A protégé plugin to support ontology axiomatization through diagramming," *ArXiv Prepr. ArXiv180810105*, 2018.

[144] M. Poveda Villalón, "Ontology Evaluation: a pitfall-based approach to ontology diagnosis," phd, E.T.S. de Ingenieros Informáticos (UPM), 2016.

[145] J. Raad and C. Cruz, "A survey on ontology evaluation methods," 2015.

[146] M. Poveda-Villalón, M. C. Suárez-Figueroa, M. Á. García-Delgado, and A. Gómez-Pérez, "OOPS! (OntOlogy Pitfall Scanner!): supporting ontology evaluation on-line," p. 9.

[147] B. Parsia, N. Matentzoglu, R. S. Gonçalves, B. Glimm, and A. Steigmiller, "The OWL Reasoner Evaluation (ORE) 2015 Competition Report," *J. Autom. Reason.*, vol. 59, no. 4, pp. 455–482, Dec. 2017, doi: 10.1007/s10817-017-9406-8.

[148] B. Glimm, I. Horrocks, B. Motik, G. Stoilos, and Z. Wang, "HermiT: An OWL 2 Reasoner," *J. Autom. Reason.*, vol. 53, no. 3, pp. 245–269, Oct. 2014, doi: 10.1007/s10817-014-9305-1.

[149] M. Horridge and S. Bechhofer, "The OWL API: A Java API for OWL ontologies," *Semantic Web*, vol. 2, no. 1, pp. 11–21, Jan. 2011.

[150] E. Blomqvist, A. Seil Sepour, and V. Presutti, "Ontology Testing - Methodology and Tool," in *Knowledge Engineering and Knowledge Management*, Berlin, Heidelberg, 2012, pp. 216–226, doi: 10.1007/978-3-642-33876-2_20.

Publications

1. Jalil Elhassouni, Abderrahim El Qadi, Yasser El Madani El Alami, Mohamed El Haziti. The implementation of credit risk scorecard using ODPs and BCBS 239. Cybernetics and Information Technologies, Volume 20, No 2, pp. 93-104 (Scopus, rang=Q2).
2. Jalil El Hassouni, Abderrahim El Qadi, Mehdi Bazzi, Mohamed El Haziti. Modeling with ontologies design patterns: credit risk scorecard as a case study. Indonesian Journal of Electrical Engineering and Computer Science Vol. 17, No. 1, January 2020, pp. 429~439 ISSN: 2502-4752, DOI: 10.11591/ijeecs.v17.i1. (Scopus, rang=Q3).
3. Jalil ElHassouni, Abderrahim El Qadi, Mehdi Bazzi, Mohamed El Haziti, Applying ontologies to data integration systems for bank credit risk management, Journal of Data Mining and Digital Humanities, ISSN 2416-5999, pp. 1-9, 2017 (DBLP).

Presentations

1. Jalil Elhassouni and Abderrahim El Qadi, Ontology Engineering Methodologies: State of the Art, BDIoT’2021, Rabat Morocco (accepted).
2. Sarah Dahir, Jalil Elhassouni, Abderrahim El Qadi and Hamid Bennis, Medical Query Expansion using Semantic Sources DBpedia and Wikidata, ISIC’2021, New Delhi, India(accepted).
3. Jalil Elhassouni, Mehdi Bazzi, Abderrahim El Qadi and Mohamed El Haziti, Applying ontologies to data integration systems for bank credit risk management, VSST’16 IEEE 2016, Rabat Morocco.
4. Jalil Elhassouni, Mehdi Bazzi, Abderrahim El Qadi and Mohamed El Haziti, Towards a data integration system based on ontologies for bank credit risk management, AIT2S’15, Settat, Morocco.

ROYAUME DU MAROC
Université Mohammed V - RABAT - جامعة محمد الخامس – الرباط –
Faculté des sciences
كلية العلوم

CENTRE D'ETUDES DOCTORALES - SCIENCES ET TECHNOLOGIES

Abstract

Recently, the amount of data generated on a daily basis has exponentially exploded. Tremendous efforts have been undertaken to provide the industry and research fields with expanding technologies for facilitating the collection, treatment, and sharing of massive data. This poses unprecedented challenges for the use of these technologies, especially in the credit risk management. The greatest challenge lies in collecting, retrieving, reusing and sharing this knowledge on one hand, and in creating new values for improving the competitiveness of financial institutions on the other hand. In the aftermath of the 2008 financial crisis, banking systems have shown a precarious fragility particularly in data architecture and information technology infrastructure. Moreover, inadequacy or non–existence of common vocabulary has created a gap in common semantics. Several solutions have been proposed, but the most efficient approach to respond to these issues is the use of ontologies. Ontologies are a natural fit for data knowledge representation, and data storing, retrieving, integrating, and reasoning. This work used the ontology for modeling the credit risk scorecard. We opted for the Krisnadhi & Hitzler methodology based on Ontology Design Patterns, which is a worked and tested example, to develop credit risk scorecard and applicant ontologies. The methodology supports collaborative construction and allows for reusability. We linked the credit risk scorecard to credit risk scorecard objectives and credit risk scoring decision support tool. Our credit risk scorecard ontology is then used to identify the corresponding tasks, subtasks, roles, actors, and resources of the affected business. Our ontology is validated using the methods and tools already tested. Firstly, we make the diagnosis and repair of the ontology for quality validation. Secondly, we reason the logical consistency of the ontology. Thirdly, we make queries answering for usability validation of the ontology. Our ontology is a shareable ontology that can be understood both by humans and computers alike. The major aim of this dissertation is to develop a rich, expandable, and re-usable credit risk scorecard based on ontology design patterns and in compliance with BCBS 239. The principal result of this work is laying the corner stone for the credit risk management platform-based ontology, which remains open to the extent that it allows for constant accumulation of tacit knowledge and efficient polymerization of explicit knowledge.

Keywords*: Ontology, Ontology Design Patterns, Knowledge-based systems, Credit risk scorecard, Credit risk management.*

Résumé

La quantité de données générées quotidiennement explose, de nos jours, à un rythme exponentiel. Des efforts immenses ont été déployés pour fournir à l'industrie et aux domaines de recherche, des technologies en expansion afin de faciliter la collecte, le traitement et le partage de données massives. Ce qui engendre des défis sans précédent quant à l'utilisation de ces technologies, particulièrement dans la gestion du risque de crédit. Le plus grand défi réside dans la collecte, la récupération, la réutilisation et le partage de ces connaissances d'une part ainsi que dans la création de nouvelles connaissances pour améliorer la compétitivité des institutions financières d'autre part. Au lendemain de la crise financière de 2008, les systèmes bancaires ont fait preuve d'une grande faiblesse notamment dans l'architecture des données et les infrastructures informatiques. En plus, l'insuffisance ou la non-existence d'un vocabulaire commun a créé une lacune dans la sémantique commune. Plusieurs solutions ont été proposées pour résoudre ces problèmes, mais l'approche basée sur l'utilisation des ontologies reste la plus efficace. Les ontologies sont une solution naturelle pour la représentation de la connaissance, le stockage, la recherche, l'intégration et l'analyse des données « reasoning data ». Ce travail a utilisé l'ontologie pour modéliser le modèle de notation « scorecard » du risque de crédit. Nous avons opté pour la méthodologie de Krisnadhi & Hitzler basée sur des patrons de conception d'ontologie « Ontology Design Patterns », une méthodologie qui a prouvé son efficacité pour le développement et le test des cas réels, afin de développer les ontologies du modèle de notation du risque de crédit et du demandeur. La méthodologie soutient la construction collaborative et permet la réutilisation. Nous avons fait la liaison du modèle de notation du risque de crédit avec ces objectifs et l'outil d'aide à la décision du risque de crédit. Notre ontologie de modèle de notation du risque de crédit est ensuite utilisée pour identifier les tâches, sous-tâches, rôles, acteurs et ressources correspondants aux domaines connexes. Notre ontologie a été validée à l'aide des méthodes et outils déjà testés. Premièrement, nous avons fait le diagnostic et la réparation pour la validation de la qualité de l'ontologie. Deuxièmement, nous avons travaillé sur le raisonnement et la cohérence logique de l'ontologie. Troisièmement, nous avons formulé des requêtes pour valider l'utilisation de l'ontologie. Notre ontologie est une ontologie partageable, compréhensive à la fois par les humains et les ordinateurs. L'objectif principal de cette thèse est de développer un modèle de notation du risque de crédit riche, extensible et réutilisable basée sur des patrons de conception d'ontologie et en conformité avec le BCBS 239. Le résultat principal de ce travail est de poser le premier bloc de la plateforme de gestion du risque de crédit basée sur l'ontologie, qui reste ouverte dans la mesure où elle permet une accumulation constante de connaissances tacites et une polymérisation efficace des connaissances explicites.

Mots-clés : *Ontologie, Patron de conception des ontologies, Systèmes à base de connaissances, Modèle de notation du risque de crédit, Risque de crédit.*

Année Universitaire : 2020/2021

✉ Faculté des Sciences, avenue Ibn Battouta, BP. 1014 RP, Rabat –Maroc
☎ 00212(0)5 37 77 18 76 🖷 00212(0)5 37 77 42 61 ; http://www.fsr.um5.ac.ma