

2015-09-30

Semantic Models as Knowledge Repositories for Data Modellers in the Financial Industry

Gary Thompson
Technological University Dublin

Follow this and additional works at: <https://arrow.tudublin.ie/scschcomdis>



Part of the [Computer Engineering Commons](#)

Recommended Citation

Thompson, G. (2015) Semantic Models as Knowledge Repositories for Data Modellers in the Financial Industry, Masters Dissertation, Technological University Dublin, 2015.

This Theses, Masters is brought to you for free and open access by the School of Computer Science at ARROW@TU Dublin. It has been accepted for inclusion in Dissertations by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie, vera.kilshaw@tudublin.ie.

Semantic Models as Knowledge Repositories for Data Modellers in the Financial Industry

Gary Thompson

A dissertation submitted in partial fulfilment of the requirements of
Dublin Institute of Technology for the degree of
M.Sc. in Computing (Information and Knowledge Management)

June 2015

DECLARATION

I certify that this dissertation which I now submit for examination for the award of MSc in Computing (Information and Knowledge Management), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Dublin Institute of Technology and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute's guidelines for ethics in research.

Signed: Gary Thompson

Date: 29th June 2015

ABSTRACT

Data modellers working in the financial industry are expected to use both technical and business knowledge to transform data into the information required to meet regulatory reporting requirements. This dissertation explores the role that semantic models such as ontologies and concept maps can play in the acquisition of financial and regulatory concepts by data modellers. While there is widespread use of semantic models in the financial industry to specify how information is exchanged between IT systems, there is limited use of these models as knowledge repositories. The objective of this research is to evaluate the use of a semantic model based knowledge repository using a combination of interviews, model implementation and experimental evaluation.

A semantic model implementation is undertaken to represent the knowledge required to understand sample banking regulatory reports. An iterative process of semantic modelling and knowledge acquisition is followed to create a representation of technical and business domain knowledge in the repository. The completed repository is made up of three concept maps hyper-linked to an ontology. An experimental evaluation of the usefulness of the repository is made by asking both expert and novice financial data modellers to answer questions that required both banking knowledge and an understating of the information in regulatory reports.

The research suggests that both novice and expert data modellers found the knowledge in the ontology and concept maps to be accessible, effective and useful. The combination of model types allowing for variations in individual styles of knowledge acquisition. The research suggests that the financial trend in the financial industry for semantic models and ontologies would benefit from knowledge management and modelling techniques.

Key Words: Knowledge Management, Semantic Model, Data Modeller, Ontology, Concept Map, Financial Information

ACKNOWLEDGEMENTS

I would like to thank the following people for their much appreciated support and encouragement and help throughout the completion of this dissertation.

Firstly, I would like to thank to my dissertation supervisor Brendan Tierney for all his guidance, advice and for always asking the relevant question.

Thanks to Deidre Lawless and Damian Gordon for their feedback on my initial research proposal and for putting the Knowledge into IKM.

Thanks also go to my colleagues at IBM for their enthusiastic participation in the interviews and experiments without which this research would not have been possible.

Finally, a massive thank you to my wife Clare for all her help, support, love and encouragement during the research and writing of this dissertation.

For Clare.

In its remote pages it is written that the animals are divided into: (a) belonging to the emperor, (b) embalmed, (c) tame, (d) sucking pigs, (e) sirens, (f) fabulous, (g) stray dogs, (h) included in the present classification, (i) frenzied, (j) innumerable, (k) drawn with a very fine camelhair brush, (l) et cetera, (m) having just broken the water pitcher, (n) that from a long way off look like flies.

– Jorge Luis Borges, *The Analytical Language of John Wilkins* (1942)

TABLE OF CONTENTS

Declaration.....	i
Abstract.....	ii
Acknowledgements.....	iii
Table of Contents.....	iv
Table of Figures.....	vii
Table of Tables	ix
1 Introduction	1
1.1 Project Introduction	1
1.2 Background	1
1.3 Research Project.....	5
1.4 Research Method	6
1.5 Research Objectives	7
1.6 Scope and Limitations.....	9
1.7 Organisation of Dissertation	10
2 Semantic Models	11
2.1 Introduction.....	11
2.2 Data Models	12
2.3 Glossaries and Vocabularies	16
2.4 Concept Maps	17
2.5 Ontologies	20
2.6 Comparison of Semantic Models.....	25
2.7 Conclusions.....	28
3 Knowledge Management and Semantic Modelling	30
3.1 Introduction.....	30

3.2	Knowledge Management and Modelling	30
3.3	Knowledge Acquisition by Modellers	33
3.4	Measures for Knowledge Acquisition and Modelling	35
3.5	Conclusions.....	38
4	Financial Information Modeller Interviews	40
4.1	Introduction.....	40
4.2	Interview Approach	40
4.3	SME Interview 1 - Models Product Manager	41
4.4	SME Interview 2 – Models Architect	44
4.5	Analysis of Interviews	47
4.6	Analysis of Contrived Knowledge Elicitation	50
4.7	Evaluation of Interviews	52
4.8	Conclusion	55
5	Implementation of Semantic Models as a Knowledge Repository	56
5.1	Introduction.....	56
5.2	Model and Tool Selection	57
5.3	Sample Reports and Data Model	58
5.4	Concept Map Implementation.....	59
5.5	Ontology Implementation	61
5.6	Evaluation of Implementation.....	65
5.7	Conclusions.....	69
6	Experimental Use of Semantic Model Based Knowledge Repository.....	70
6.1	Introduction.....	70
6.2	Experimental Approach	70
6.3	Design of Questions and Metrics	72
6.4	Overview of Experimental Sessions and Participants	75

6.5	Results from Experiment Questions	76
6.6	Results from Feedback Questions.....	83
6.7	Evaluation of Use of Knowledge Repository	86
6.8	Conclusions.....	90
7	Conclusions and Future Work.....	92
7.1	Introduction.....	92
7.2	Research Definition & Research Overview	92
7.3	Contributions to the Body of Knowledge	94
7.4	Experimentation, Evaluation and Limitation.....	96
7.5	Future Work and Research.....	97
	Bibliography	98
A.	SME Interview Questions.....	103
B.	Interview Knowledge Analysis.....	105
C.	Triadic and Card Sort for SMEs	106
D.	Experiment Logical Data Model	110
E.	Experiment Concept Maps	111
F.	Experiment Ontology Term List	114
G.	List of Supporting Material	117

TABLE OF FIGURES

Figure 1.1: Extract from AIB 2013 Pillar 3 Disclosure (AIB, 2013)	2
Figure 1.2: Examples of Business and Information Questions	2
Figure 1.3: Types of Knowledge and Semantic Models in Research Project	5
Figure 1.4: Project Approach and Activities	6
Figure 2.1 : Spectrum of semantic models (Bergman, 2007)	11
Figure 2.2 : Example of Logical Data Model (Moody and Shanks, 2003)	13
Figure 2.3 : Example Dimensional Logical Model (Kimball and Ross, 2011)	14
Figure 2.4 : Data Model Quality Factors (Moody and Shanks, 2003)	15
Figure 2.5 : A Simple Concept Map	18
Figure 2.6 : Example Concept Map for Accounting Theory (Simon, 2007)	19
Figure 2.7 : Extract from Wine Ontology (Noy and McGuinness, 2001)	20
Figure 2.8 : RDF Triple pseudo-code and Graph Visualization (W3C, 2014).....	22
Figure 2.9 : Structure of OWL (W3C, 2012).....	23
Figure 2.10 : Subset of FIBO Business Entity Diagram (EDM Council, 2014)	24
Figure 2.11 : Dimensions of an Ontology Model (Benne, 2014)	25
Figure 2.12 : Ontology as Conceptual Model for Data (Bennet, 2014)	25
Figure 2.13 : Mapping C-map to Ontology (Graudina and Grundspenkis, 2011).....	27
Figure 2.14 : Mapping of Terms and Concepts (Obrst, 2008).....	27
Figure 2.15 : Where Definitions are Found (Adapted from Chisholm 2010)	28
Figure 2.16 : Sematic Models Types in Chapter 2	29
Figure 3.1 : The SECI Process (Nonaka <i>et al.</i> , 2000).....	32
Figure 3.2 : Screenshot from E-Learning Environment (Vieritz <i>et al.</i> , 2013).....	36
Figure 4.1 : Concept Model based on Contrived Knowledge Elicitation.....	51
Figure 5.1 : Iterative Process of Model Implementation	56

Figure 5.2 : Concept Map for Basel II Capital Adequacy Framework.....	59
Figure 5.3 : Ontology Term for Tier 1 Capital	61
Figure 5.4 : Equivalency of Hierarchies in Ontology and Concept Map	63
Figure 5.5 : Result of Searching Ontology for Capital Ratio	64
Figure 5.6 : Example of Relating an Ontology Term to Data Model Attribute	65
Figure 5.7 : Human Consumable and Machine Implementable Knowledge.....	68
Figure 6.1 : Example of Experimental Questions.....	71
Figure 6.2 : Example of Format Used in Experiment Questions.....	74
Figure 6.3 : Photo of Participant Performing the Experiment Tasks.....	75
Figure 6.4 : Distribution of Most Useful Model for All Questions	77
Figure 6.5 : Distribution of Helpful Rating for All Questions	78
Figure 6.6 : Analysis of Question Correctness by Expert and Novices.....	79
Figure 6.7 : Distribution of Model Usefulness by Question Category	80
Figure 6.8 : Distribution of Model Usefulness by Participant.....	81
Figure 6.9 : Distribution of Answer Correctness by Most Useful Model	81
Figure 6.10 : Distribution of Helpful Rating by Question Category	82
Figure 6.11 : Distribution of Helpful Rating by Most Useful Model.....	83
Figure 6.12 : Distribution of Answer Correctness by Helpful Rating.....	83
Figure C.1 : Concepts selected for Card Sort and Triadic.....	106
Figure C.2 : Results of First Card Sort by Subject Matter Experts	107
Figure C.3 : Results of Second Card Sort by Subject Matter Experts	108
Figure D.1 : Logical Data Model from Experiment	110
Figure E.1 : Basel II Capital Adequacy Framework Concept Map.....	111
Figure E.2 : Credit Risk Concept Map	112
Figure E.3 : Capital Requirements Concept Map.....	113

TABLE OF TABLES

Table 1.1 : Project Research Questions and Evaluation	8
Table 2.1 : Examples of Glossary Terms (Chisholm, 2010)	16
Table 2.2 : Mapping of Physical Data Model to Ontology Elements.....	26
Table 3.1 : Models or Representation of Types of Knowledge.....	31
Table 5.1 : Sample Reports and corresponding Data Model Fact Entities.....	58
Table 6.1 : Comparing Business and Information Questions	72
Table 6.2 : Experiment Questions	73
Table 6.3 : Metrics for Model Usefulness	74
Table 6.4 : Profile of Experiment Participants	76
Table 6.5 : Summary of Experimental Question Metrics	78
Table 6.6 : Helpful Rating Summary Statistics	82
Table B.1 : Occurrences of Concepts in Interview	105
Table C.1 : Triadic Results	109
Table F.1 : List of Ontology Terms	114
Table G.1 : List of Supporting Material.	117

1 INTRODUCTION

1.1 Project Introduction

Data modellers working in the financial industry are expected to use technical and business knowledge to transform data into the information that can be presented in operational and regulatory reports. Ontologies are used in the financial industry to provide the specification for the sharing of information between information systems. While these ontologies provide a formal structure of financial information, they impart limited conceptual understanding to end-users on how the information should be interpreted or used. This contrasts with the use of concept maps in financial education to support students in their acquisition of business knowledge on topics such as advanced accounting concepts.

This project evaluates the use of an ontology and concept maps to encapsulate the knowledge required to model and understand the information presented in a set of financial regulatory reports. Expert and novice financial data modellers were asked to perform tasks that required both business domain knowledge and an understanding of the information structure of the report. An evaluation is made of the use of the semantic models to measure their usefulness in completing the tasks and to assess how concept maps and ontologies can complement each other in a knowledge repository. The results of the experiment are compared with the similar studies in the literature and with the views of experienced data modellers.

1.2 Background

Data modellers require knowledge as they design the data models that underpin the information solutions created to meet business requirements. Examples of design tasks include the selection of data model attributes for a business report or the addition of a new attributes to meet a regulatory requirement. To perform these tasks the data modeller requires an understanding of both how report will be used by business users and of the structure of the information stored in databases. This knowledge requirement is succinctly summarized by Smith (2003) in what he refers to as the Otologist's Credo : 'To create effective representations it is an advantage if one knows something about the things and processes one is trying to represent.'

A financial regulatory report may not always have a complex structure but it will contain information that requires an understanding of specialist jargon and financial domain knowledge to model and consume the information. The Basel Pillar 3 Compliance Adequacy Disclosures are an example of reports published by banks that are based on financial industry regulatory reporting standards. An extract from the 2013 Pillar 3 Disclosure (AIB, 2013) from Allied Irish Bank is provided in Figure 1.1.

	2013	Restated* 2012
	€ m	€ m
Total capital	10,336	12,682
Risk weighted assets		
Credit risk	59,038	66,335
Market risk	177	616
Operational risk	3,180	4,466
Total risk weighted assets	62,395	71,417
Capital ratios		
Core tier 1	14.3%	15.2%
Total	16.6%	17.8%

Figure 1.1: Extract from AIB 2013 Pillar 3 Disclosure (AIB, 2013)

A data modeller developing a model to support a business report will have to ask questions to ensure they have a correct understanding of the business information that is to be supplied to users. Examples of the types of questions asked data modellers and business users are illustrated in Figure 1.2.

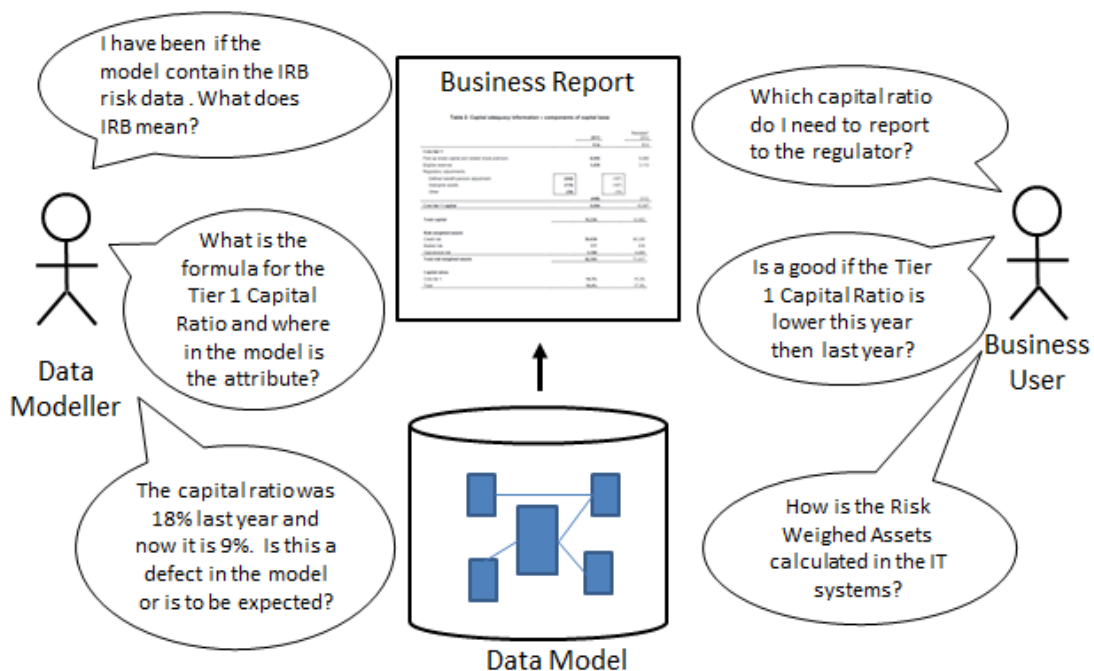


Figure 1.2: Examples of Business and Information Questions

Chisholm (2010) is of the opinion that the data represented by a data model can only be understood and turned into information if there are adequate definitions that provide a shared semantic reference for the information creator and consumer. Semantic models such as ontologies and concept maps are related to relational data models and the Universal Modelling Language (UML) used in information management (Blaha, 2010). Given the volume and variety of structured and unstructured data available to banks and other financial institutions there is a need for semantic models that are both human and machine readable (O'Sullivan *et al.*, 2014). An example of the semantic challenges faced by the industry is the adoption of predictive and data mining analytics for risk modelling that is resulting in an increased complexity in specialist business and technical terminology (Zhu and Huang, 2014).

Abi-Lahoud *et al.* (2014) describe a transaction problem that typically occurs between semantic technical experts and business subject matter experts during the development of semantic models. They identified that this has contributed to poor financial regulation and governance due to failures in translating business requirements into implemented IT systems. Their research show that the translation problem can be overcome through the use of a formal language called Structured Business Vocabulary and Rules (SBVR) to act as a means of common communication between technology and business. They found the business subject matter experts in the financial domain were comfortable expressing their knowledge using a formal language, which in turn can be used by the technical experts to implement semantic models of IT systems.

A challenge to creating semantic models is that there is no consensus on the format or meta-model for model development (Rodriguez-Priego *et al.*, 2010). Common information interchange languages such as eXtensible Business Reporting Language (XBRL) and SBVR have been used for communication between national supervisory authorities and financial entities since the mid-2000s (Bonson-Ponte *et al.*, 2007)(Abi-Lahoud *et al.* (2014). The Financial Industry Business Ontology (FIBO) is an being undertaken by the Enterprise Data Management Council (EDM Council) with to formally specify financial terms and definitions using OWL and UML. The purpose of FIBO “is for data harmonization and for the unambiguous sharing of meaning across data repositories. This common language (or Rosetta stone) for the financial industry supports business process automation and facilitates risk analysis” (EDM Council, 2014).

Concept maps are used to capture and communicate knowledge as part of a knowledge management projects and have been successfully used to assist in the teaching of financial and technical concepts (Simon, 2007; Vieritz *et al.*, 2013). Ontologies emphasize formal rules for how concepts can be related to each other and are commonly developed to govern the sharing information between information systems (Correndo and Alani, 2007; De Vergara *et al.*, 2004). Less formal semantic model types such as structured vocabularies can be useful in the initial stages of ontology development (Abi-Lahoud et al, 2014).

The discipline of knowledge management concerns itself with the creation, communication, management and effective use of knowledge by humans. Knowledge management supports these activities within the three main themes of people, process and technology. Pope and Butler's (2012) review of knowledge management systems concludes that implementations must take a holistic approach that encompasses all three of these themes. The data, information, knowledge and wisdom (DIKW) pyramid model is widely used within the knowledge management literature to encapsulate how technology system transforms data into actionable information that can be acted on by humans (Rowley, 2007). Rowley (2007) states that DIKW pyramid as articulated by Ackoff (1989) emphasises that algorithms implemented programmatically in IT solutions are suited for the processing of data, but are less suited for the processing of knowledge. This early expression of the translation problem shows that challenge of turning business knowledge into implementable code has been an issue for both knowledge management and information management for over twenty five years.

Data modellers are semantic technical experts who encounter the translation problem while working in industries such as banking and finance where with business requirements are expressed in complex sector specific jargon. While formal ontologies such as FIBO could be considered as a repository of deep explicit knowledge on the structure and composition of financial information, a less formal model such as concept maps or structured vocabularies could be more useful to data modellers when acquiring tacit knowledge of financial and regulatory concepts. The use of ontologies and concept maps can be seen as complementary – the applicability of either to a given task being determined by the activity, individual expertise and individual preferences.

1.3 Research Project

The motivation for this project is to investigate combinations of knowledge and semantic modelling techniques that can be applied to make the data modeller live easier when acquiring business domain knowledge. Ontologies and logical data models created by technical modellers are viewed as technical artefacts whose primary purpose is to support the development of IT solutions used by business users. They are not regarded as repositories of business knowledge to be used by data modellers. The literature identifies that concept and ontologies are part of a spectrum of models that contains data model, structured languages and glossaries of terms (Obrst, 2003). While there is discussion in the literature on how the structural elements of concept maps, ontologies and data model can be mapped to each other, there is limited discussion on how they can be used in combination to assist knowledge acquisition by data modellers (Graudina and Grundspenki, 2011; Osman *et al.*, 2011).

The hypothesis of this research project is that data modellers can usefully learn knowledge when it is represented in an integrated set of ontology and concept maps. Specifically that a data modeller with no experience working in the financial industry is able to understand and interpret the information in Basel 3 Disclosure Report and associated data model. The data modeller requires conceptual knowledge about financial regulation and reporting structures, and procedural knowledge about the calculation of financial information. The types of information, knowledge and semantic models included in the research are illustrated in Figure 1.3.

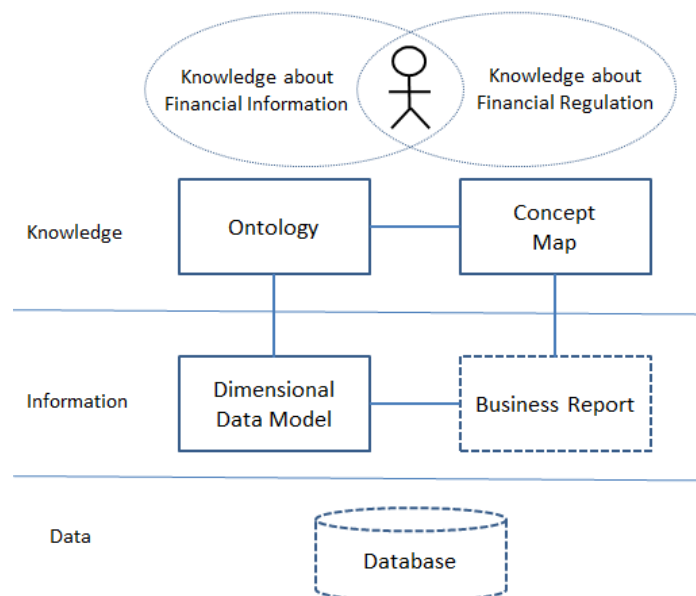


Figure 1.3: Types of Knowledge and Semantic Models in Research Project

1.4 Research Method

The approach to address the project objectives is divided into the three stages of background research, implement and experiment illustrated in Figure 1.4. This approach was chosen as it allows for an exploration of the knowledge management themes of people, process and technology. The opinions and experiences of data modellers are captured during both background research and experimentation. The implementation requires the use of knowledge acquisition and modelling processes to represent the knowledge in the semantic models. The experiment will evaluate the technical implementation of the models as it is used by the data modellers.

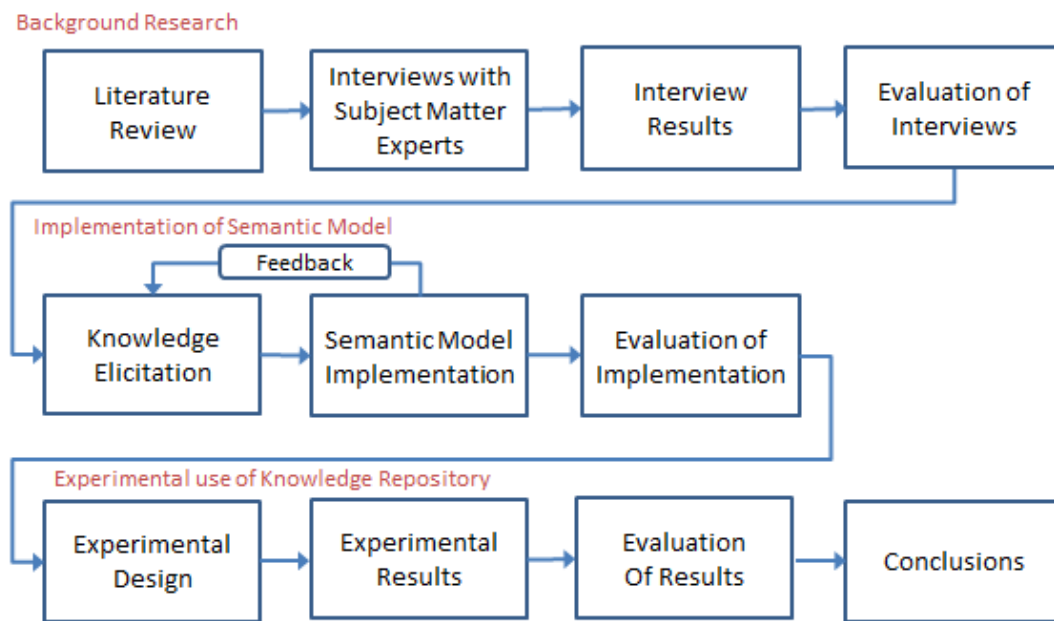


Figure 1.4: Project Approach and Activities

A review of the literature will be conducted covering the areas of semantic models, data modelling and knowledge management. This will include a comparison of the different styles of semantic models used in the financial industry. Interviews will be conducted with financial data modellers to obtain their opinions on the use of semantic models in the industry and the role semantic models play in data modeller knowledge acquisition. The literature review and the interviews are required to evaluate the variety of styles, languages and tooling formats for the ontology and concept maps to be used in the research.

A knowledge repository comprising of an ontology and concept maps will be implemented to represent the knowledge required to understand the business report and

its associated dimensional model. Knowledge elicitation will be conducted using regulatory standard documents, regulatory reports created by Irish banks and reviews with expert financial data modellers. The project will use published Basel Pillar 3 Compliance Adequacy reports to obtain sample reports and to provide examples of business concepts to be modelled. These reports have been chosen as they are a financial industry standard and there are examples are publicly available from Irish banks. An evaluation will compare the experiences observed during this implementation with views from both the literature and subject matter expert interviews.

The implemented knowledge repository will then be used by data modellers when answering questions about the information in the financial reports and an associated data model. The participant's use of the semantic models will be measured using both quantitative and qualitative methods. The experiment will provide results on the effective use of the ontology and concept maps by the participants performing pre-defined tasks. Quantitative measures will be gathered through the use of a questionnaire and qualitative measures will be gathered using observations and comments made by user as they conduct tasks. An evaluation will be made by comparing the use of ontologies and semantic models as knowledge bases, their relative relevance to expert and novice users, and the influence software tooling choice in the implementation. Comparisons will be made and discussed with respect to academic and industry literature relevant to semantic modelling and knowledge management.

1.5 Research Objectives

The overall objective of this research is to evaluate the use of the semantic models as knowledge repository that is useful to data modellers working with financial regulatory reports. The individual objectives of the research project are;

- Review existing academic and industry literature to identify the use of semantic models and concept maps used in financial information management.
- Investigate use of the semantic models by expert financial data modellers.
- Implemented a semantic model based knowledge repository to observe modelling and knowledge management challenges.
- Design and execute an experiment to assess the relative merits and disadvantages of using ontologies and concepts maps to represent business and

technical knowledge required by data modellers working in financial services industry.

- Document and evaluate the findings from the experiment.
- Compare the results of the experiment with the current views in the literature and suggest how semantic models could be better used for financial information modelling.
- Make recommendations for any future research in this area.

To meet these objective the project will apply the relevant research methods to address and evaluate the eight research questions presented in Table 1.1.

Table 1.1 : Project Research Questions and Evaluation

#	Research Question	Evaluation Method
Semantic and data models in the financial industry		
1	What types of semantic models are used in the financial industry?	Contrast subject matter expert interviews with literature review
2	Are semantic models currently used by data modellers to acquire knowledge?	
Implementing semantic models as knowledge repository		
3	What are the considerations when implementing semantic models as a knowledge repository?	Discuss implementation experience Contrast with interviews and literature review
4	What is the difference between semantic modelling and knowledge modelling?	
Using Semantic models as a knowledge repository		
5	Do semantic models provide useful knowledge to data modellers performing tasks related to a financial regulatory report?	Discusses and compare experiment quantitative results and participant comments. Compare with similar studies in literature review. Compare results with expectations from implementation
6	Do semantic models better represent knowledge about the business domain or the structure of the information?	
7	Is a concept map or an ontology more useful in the knowledge repository?	
8	Are the semantic models more useful to novice or expert data modellers?	

1.6 Scope and Limitations

The scope of this research is to evaluate the use of an ontology and concepts maps with a group of data modellers who are all part of the same geographically located work group. The study requires the participation of individuals that have a minimum of the one year of applied data modelling experience. The organisation where the study will be conducted includes a group of over twenty data modellers with a mix of financial and non-financial modelling experience.

The semantic models developed for this project will be fully functional in the selected tool, but it is not possible to deploy the models into the work environment to observe how they could be used on a day to day basis. The limitations were due to the unavailability of the participants to partake in a longer study. The restricted access to the experts also limited the involvement of five data modellers in the detailed evaluation.

The purpose of this research is to create representative but not an exhaustive set of semantic models to represent business domain knowledge related to Basel capital requirements and credit risk management. A small number of business reports will be selected as a sample of the regulatory reports that are made available by the regulated bank in Ireland. The scope of knowledge in the repository is also limited by the time available to the author to both acquire and model the required knowledge.

It is not an objective of this research to evaluate the relative use of the wide range of the semantic model types and modelling tools. The implementation is limited to ontologies and concept maps as representative examples of formal and informal semantic models. The selection of the tools was limited those likely to be most usable by the participant as the objective of research was to evaluate the knowledge model content more than the tool functionality. Existing financial semantic models were not included as the independent development of the models is considered an important aspect of the research.

1.7 Organisation of Dissertation

Chapter 2 introduces the discipline of semantic modelling and describe the characteristic of different model types including concepts maps, ontologies and data models. It examines how the different models types are commonly viewed as being part of spectrum of semantic complexity and also how the structures of one of model type can related to others.

Chapter 3 examines the overlap between the disciplines of knowledge management and semantic modelling. It compares the process of knowledge acquisition as described in knowledge management with the processed of semantic modelling and technical modeller expertise development. The chapter also introduces the metrics and experimental approach used in previous research on modelling and knowledge acquisition

Chapter 4 summaries and evaluates two interviews conducted with experienced financial information modellers on the use of semantic models in the financial industry and their own experiences in acquiring technical and business data modelling knowledge. The analysis of the interview compliments the background research conducted in Chapters 2 and 3.

Chapter 5 describes the creation and implementation of an ontology and concept models as a knowledge repository for use in the research experiment. The technical tooling and modelling challenges encountered during the modelling process are identified and discussed.

Chapter 6 presents the design and results of experimental use of the ontology and concept models implementation by data modellers who were asked question relating to sample financial regulatory reports. The section included a detailed analysis of the metrics gathered during the experiment and how these results compare with other similar studies.

Chapter 7 presents the conclusion of this research, how it has contributed the body of knowledge and suggestions for further research.

Appendices A to G contain material that supports the discussion in the dissertation. Additional supporting material including interview transcripts, detailed result data and semantic models files are included in the electronic documents that accompany this dissertation document.

2 SEMANTIC MODELS

2.1 Introduction

This chapter provides an overview of a variety of data, information and knowledge modelling techniques that describe themselves as being involved in the creation of semantic models. Sections 2.2 to 2.5 present a review of semantic models types including data models, vocabularies, concepts maps and ontologies. Concept maps and glossaries are types of semantic models that are used to capture knowledge as part of a knowledge management project (Milton, 2007). Ontologies were initially developed within the fields of artificial intelligence and natural language processing and are now used to support the sharing of knowledge by both machines and humans in wide number of business and academic domains (Fensel, 2003). In database design, a logical data model is referred to as semantic model when compared to a physical data model used to implement the data structures (Zachman, 1987), (Angles, 2012).

Obrst (2003) suggests that the different types of semantic models can be compared to each other in a semantic spectrum that ranks models using factors such as semantic explicitness and complexity of implementation. While a taxonomy allows for the hierarchical classification of concepts it lacks the expressiveness of an ontology to identify the attributes of concepts or relationships between concepts (Kramar, 2013). Ontologies provide a rich set of formal representations allowing them to be used to create both machine-interpretable semantic models and to develop knowledge bases.

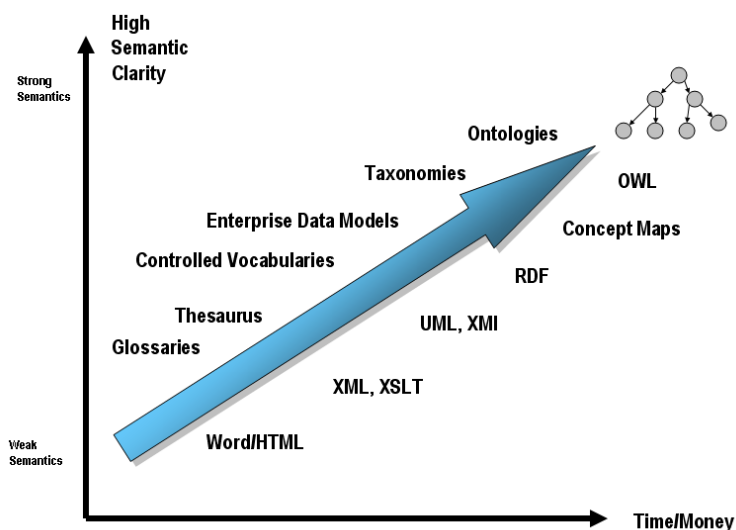


Figure 2.1 : Spectrum of semantic models (Bergman, 2007)

The spectrum in Figure 2.1 compares the model types in term of semantics strength - expressiveness or representational formality, with Time/Money - a pragmatic measure of the effort to implement and use the model types. Moody and Shanks (2003) suggest that pragmatic aspects such as measuring the effectiveness of model implementation or evaluating the quality of the model from an end-user perspective is often discussed theoretically but is not frequently empirically validated

Data architectures traditionally built on relational database systems are being augmented with a variety of data persistence technologies such as Hadoop, graph databases, columnar databases, document database and key-value stores (Kimball, 2011), (Sadalage and Fowler, 2013). The variety and volume of data stored in such architectures is increasing rapidly and requires strong data governance if information is to be extracted usefully and accurately for tasks such as financial regulatory reporting (Malik, 2013). The extraction of meaning from the combination of unstructured or semi-structured data involves the use of semantic models. Financial regulators expect banks to have the capability accurately and reliably aggregate, validate and reconcile risk data to provide an authoritative single source of risk information. The Basel Committee on Banking Supervision (2013) state that a precondition of this capability is common dictionary of risk concepts that supports the consistent use of information across the banks automated and manual systems). There is a need in the financial industry for semantic models that aid the implementation of information systems and to support the activities of financial regulators, business users and technical specialists (Abi-Lahoud *et al.*, 2014).

2.2 Data Models

Data models are created by data modellers to represent the structure of a database to ensure that the implemented database can support the information requirements of the end user. The levels of data model abstraction described by Zachman nearly thirty year ago are still applied in enterprise data model design (O'Sullivan *et al.*, 2014). The three levels of the Zachman (1987) framework are;;

- Conceptual data models that identify the data entities and their relationships at a high level understand by the business user. It identifies the data that is required by end users and sets the scope for further modelling tasks.

- Logical data models which are a formal specification of the structure or schema that is created and used by the data modeller. It specifies the details of the data elements, the data types and allowed relationships in the data set
- Physical data models which are the implementation of logical models to a specific database technology. It defines the database structures that manage the data such as allocation to disk memory and indexes for query access. This is the level that is used by the developer building the database.

Logical data models (LDM) are created during the design stage of the IT solution lifecycle where the data modeller identifies the data elements and relationship required to meet the functional use cases of the solution. Data modelling is one of the most critical tasks that influence the quality of information solutions as any defects introduced at these stages will be more expensive to resolve at a later point (Moody and Shanks, 2003). The resulting LDM is a logical database schema that is independent of a specific database technology but it is used as the primary input for the creation of the technology specific database. (Kuper and Vardi, 1993). The LDM is the bridge between the real-world or business meaning of the data and the technical implementation of the database. It can be viewed as putting structure on the data as so that it becomes information that is useful to end user per Ackoff's DIKW Pyramid (Rowley, 2007). An example of a logical data model is shown in Figure 2.2.

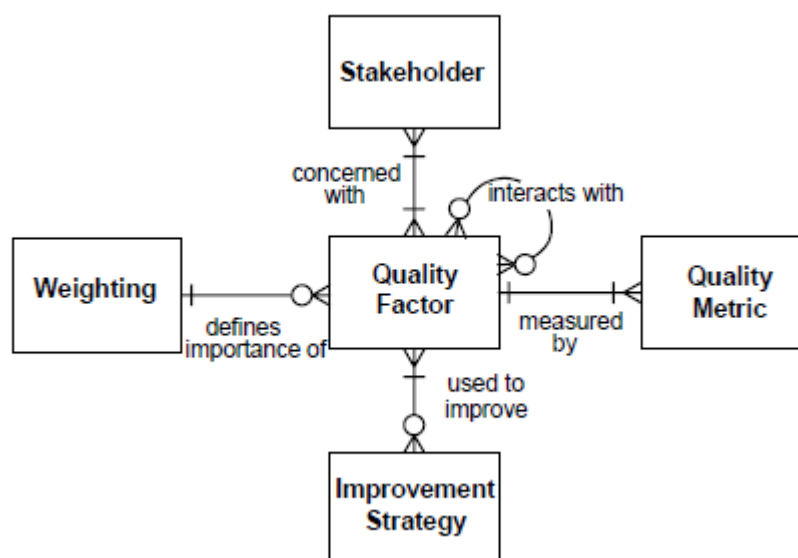


Figure 2.2 : Example of Logical Data Model (Moody and Shanks, 2003)

A logical model for an RDBMS is primarily expressed using the relational notation entities, attributes, keys and relationships proposed by Codd in the 1970's (Blaha, 2010).

- Entities are an abstraction of a type of object or event about which data is being stored. Each entity in the model is uniquely identified by a name in the model. An entity represents all possible instances of the object being stored in the database.
- Attributes are the pieces of data to be stored for each instance of an entity, for example a person's name, their date of birth and their customer number.
- Relationships are connections between entities that how instances one entity can be related to instances of other entities e.g. each customer must have one or more postal addresses.
- The primary key of entity that uniquely identifies each instance of an entity.
- Foreign keys are attributes of an entity that are primary keys of related entities.

Two prominent styles of entity relationship modelling are Third Normal Form (3NF) and Dimensional. Ralph Kimball is a proponent of dimensional modelling who suggest that the highly normalised structures of 3NF are difficult for both data modellers and end-users to use (Kimball and Ross, 2011). Dimensional models collects business metrics that are commonly used together into fact entities so that the data can easily analysed by users. The fact entity is related to dimensional entities that define the different ways in which the measures in the fact can be analysed, as shown in the example in Figure 2.3.

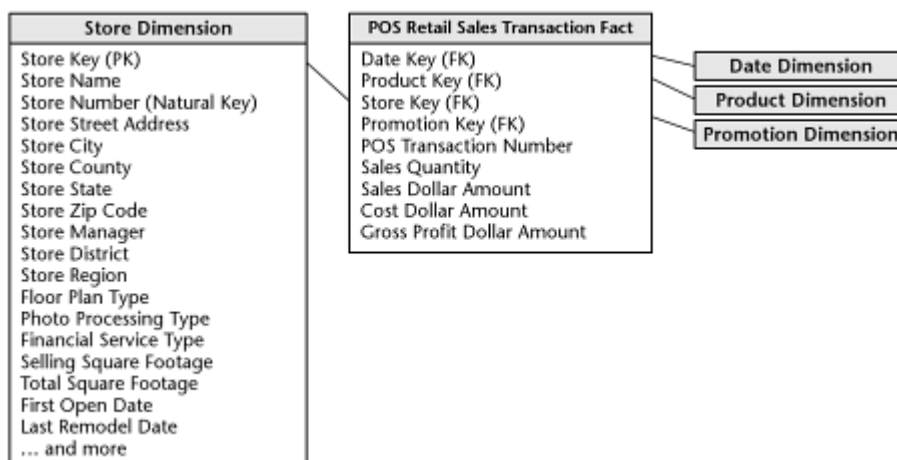


Figure 2.3 : Example Dimensional Logical Model (Kimball and Ross, 2011)

Moody and Shanks (2003) have proposed and validated a quality review framework that can be applied to the development of conceptual and logical data models. The quality factors emphasise the importance of communication between the stakeholders required for data model developed:

- Business Users who define the requirements of information and are the ultimate end users of the information in the solution
- Data Analyst or Modeller who develops the conceptual and logical data models
- Database Administrator who integrating it into the organisations information management architecture
- Application Developers who implement the physical data model

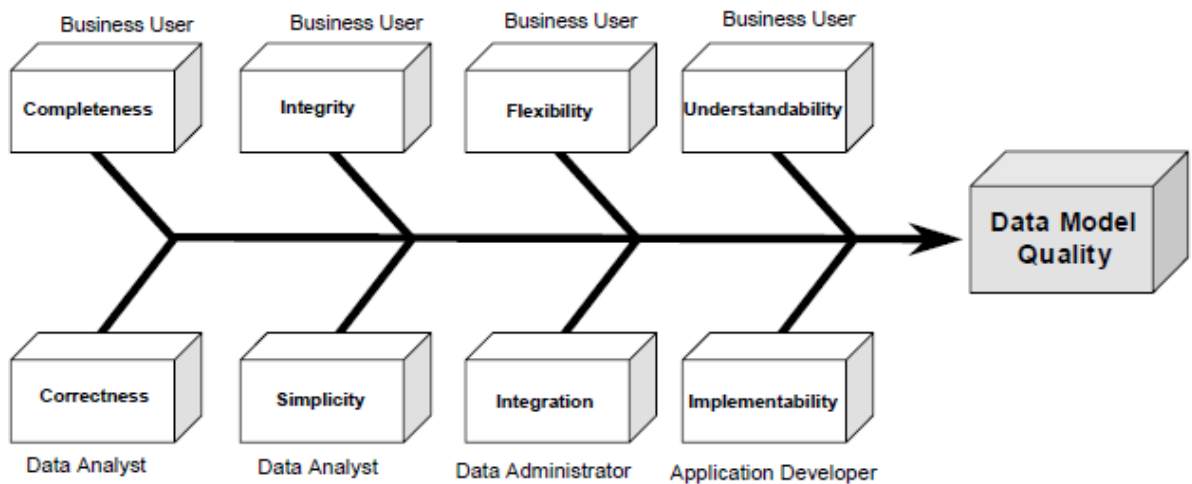


Figure 2.4 : Data Model Quality Factors (Moody and Shanks, 2003)

The eight quality factors in the framework are illustrated in Figure 2.4 and are described as follows;

1. *Completeness.* The data model contains all of the information requirements of the business users.
2. *Integrity.* The data model reflects and supports the enforcement of business data rules. This is especially important in the financial industry where there is need to guarantee data integrity and to enforce regulatory policies.
3. *Flexibility.* The data model easily accommodate changes in the business or regulatory requirements without e.g. the addition of a new product, new reporting requirement

4. *Understandability*. The data structure and to be understood by a business user who is discovering the data available in the information solution.
5. *Correctness*. The data model accurately and efficiently implements models constructs such as normalisation, sub-typing, and mandatory fields.
6. *Simplicity*. The data model contains the practically fewest number of entities and attributes to meet the requirements and is not over complicated.
7. *Integration*. The data model fit into the enterprise data model or information architecture.
8. *Implementability*. The logical data model can be realistically implemented in as physical data model.

2.3 Glossaries and Vocabularies

A glossary is a list of terms and definitions and is considered a basic type of semantic model that lacks the expressiveness of other model types (Gruninger *et al.*, 2008). Glossaries provide a reference for end-user to understand domain-specific jargon and terminologies that is used by subject matter experts, and sn individual who is familiar with the definition or jargon of a particular industry is seen as someone who ‘knows the business’ (Chisholm, 2010). The glossary may comprise of textual descriptions, acronyms, synonyms of terms.

Table 2.1 : Examples of Glossary Terms (Chisholm, 2010)

Term	Definition
Database	A store of related data.
Data Model	A technique, usually diagrammatic, used to design or document the structure of data.
Field	An instance of a data structure in which a single value can be placed. A record is said to have fields. Fields can occur in data structures that are not necessarily relational databases.

Chisholm (2010) states that in the context of business data management, the words and descriptions in the glossary are definitions that should provide end-users with a better understanding of concepts represented in information systems. A word or term in a glossary is simply a label used in natural language to communicate concepts. This can

lead to challenges when two individuals use the same word to refer to different concepts. Chisholm (2010) defines describes a good definition as a *real definition* that fully explains a concept rather than a *nominal definition* that describes just the term or simply identifies an instance. His description uses specific definitions of *term*, *concept* and *instance*;

- A *Term* is a written or verbal label that is used in language to represent a concept. It is used to both refer to a concept or the instances of a concept.
- A *Concept* is an abstract or generic idea inferred from specific instances. It can be considered analogous with entities in data models.
- An *Instance* is an individual physical implementation of a concept or the representation of a physical instance.

The Semantics of Business Vocabulary and Rules (SBVR) is a standard for the development of vocabularies that is developed by the Object Management Group (2015). The SBVR provides a method of consistently documenting the semantics of business vocabularies and rules to facilities their exchange and use between organisation and information systems. Abi-Lahoud et al (2014) description the use of SBVR to a create definition for financial regulation emphasise those parts of the SBVR specification that are similar to the definitions used by Chisholm.

- An *Expression* is a sound, text or gesture used to communicate a concept or meaning, where one expression can be associated with multiple concepts.
- A *Noun Concepts* is a group of things of interest in the domain of the vocabulary, for example banks or financial regulators
- An *Individual Noun Concept* represent actual instances of Noun Concepts, for example Wells Fargo Bank (bank) and Securities and Exchange Commission (regulator)

2.4 Concept Maps

Concepts Maps are knowledge modelling tool that was developed by Joseph Novak of Cornell University as part of research in the children's knowledge of scientific knowledge. The theoretic basis for concept maps is derived from the cognitive psychology view that individuals learn new concepts by fitting them into their mental model that made up previously understood concepts and propositions (Novak and Cañas, 2006). Concept maps have been used as a visual aid to assist in the teaching of

financial and accounting concepts (Greenberg and Wilner, 2015). The educational benefit of concept mapping is to due increased leaner participation when compared listening or reading text (Nesbit and Adesope, 2006).

Concept maps are visual models made up of concepts and propositions. Concepts are represented as a text box or bubble that contains words or symbols, as illustrated in the example in Figure 2.5. Novak and Cañas (2006) defines concepts as '*perceived regularity in events or objects, or records of events or objects designated by a label*'. Concepts are connected by lines that are labelled with words to form proposition which are the combination of two or more related concepts to make a meaningful statement about an object or event.

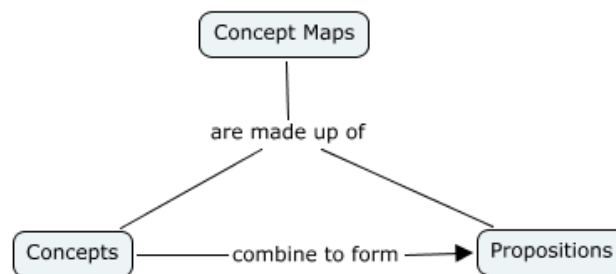


Figure 2.5 : A Simple Concept Map

Novak and Cañas (2006) state that the process of constructing of a good concept maps begins with an individual who is familiar with the domain of knowledge and that there is clear focus or question for the map. This starting point is very similar to Smith’s (2003) Otologist’s Credo that to develop a good ontology it is beneficial that the otologist has some understanding of what is being represented. The context of the concept map may be provided by a piece of text, a problem statement or an existing disparate set of knowledge artefacts to be related to each other. The key concepts are then identified and ranked either by the individual modeller or collaboratively by a group. The identification of concepts may use both natural and contrived knowledge elicitation such as cards sorts and document reviews.

Once the concepts have been identified, an initial development of the concept map can be made either with specialist concept mapping software or simpler techniques using post-it notes or whiteboards (Milton, 2007). Whatever the technology used, it should allow for iterative modification with the recognition that the concept map will never be completed as is always subject to revision. Novak and Cañas (2006) identify that

organisation of concepts into a hierarchical structure can be difficult for individuals new to concepts maps. The initial modelling leads to the identification of the cross-links between concepts in different part of the map. Discussion of the cross-links leads to clarification and reordering of existing relationships. A syntax check of the completed concept map is then undertaken to ensure that the concepts and propositions can be read as meaningful sentences. An example of cross-links is show in the accountancy education themed concept maps in Figure 2.6.

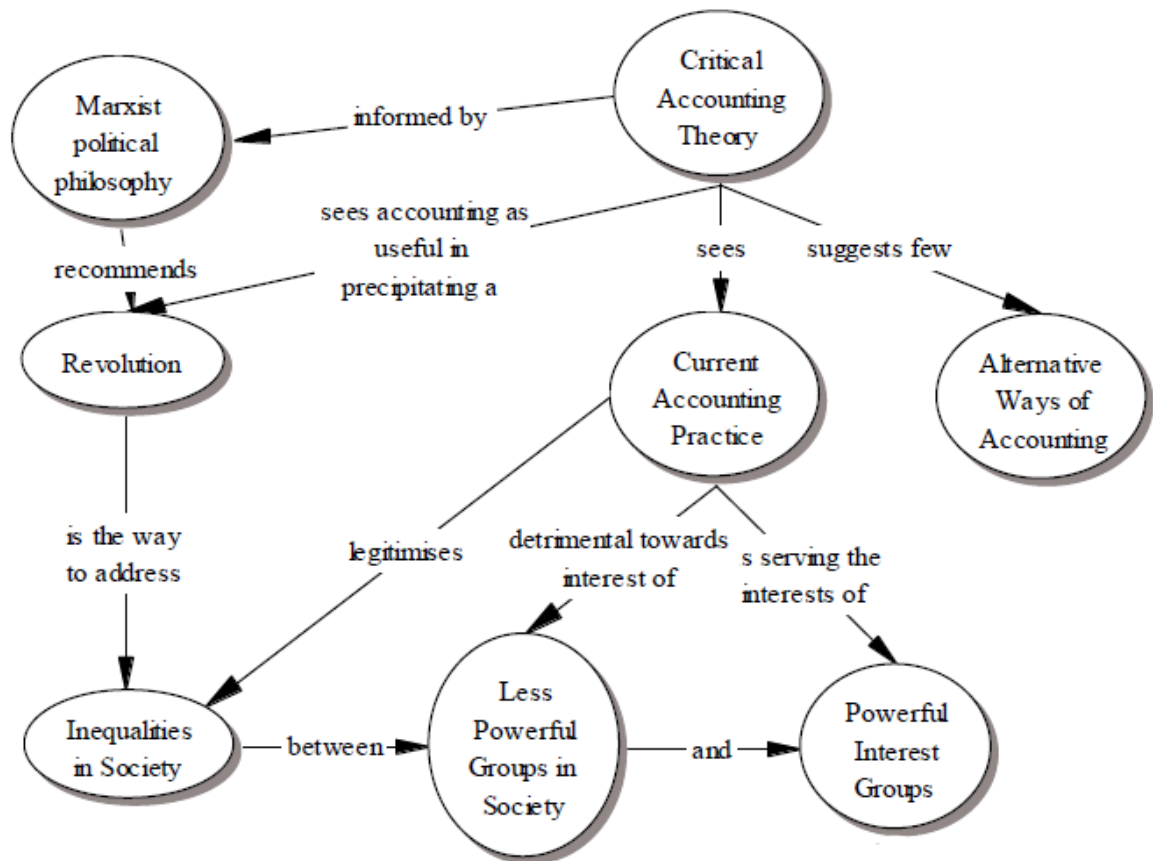


Figure 2.6 : Example Concept Map for Accounting Theory (Simon, 2007)

There is no consensus on the metrics used to define a good concept map. Milton's (2007) opinion is that while there is no right or wrong way to create a knowledge model such as a concept, it is better to have a small number model rather than one large model. Each of these models should contain a restricted number of metrics and be neat with no crossing links. Åhlberg's (2013) review of the guidance given in the literature identified a variety but no consistent opinion on the use of arrows in cross-links, the use of short or long phrases in propositions, representation of concepts with images, and preference for hierarchical or non-hierarchical arrangements

2.5 Ontologies

An ontology is an explicit and unambiguous specification of the common words and concepts used to describe and represent a domain of knowledge (Albarrak and Sibley, 2009). Obrst (2003) describes an ontology as comprising of classes of things or concepts in the domain; the relationship between classes, the properties of classes; the functions involving the classes and the rules or constraints applied to the classes. An extract from an ontology showing classes and relationships is represented visually in Figure 2.7. Ontologies are expressed in using formal languages and software tools that support a consistent and accurate specification of the concepts. The development of an ontology requires both expertise in the domain being modelled and skills in the use of ontology modelling tools, which typically involves an ontology modeller working with a subject matter expert in what can be a long running modelling process (Albarrak and Sibley, 2009). Their formal structure makes ontologies suitable for the development of knowledge bases that are reusable and machine readable, with the knowledge base being created when an ontology is combined with instances of the concepts (Noy and McGuinness, 2001), (Obrst, 2003).

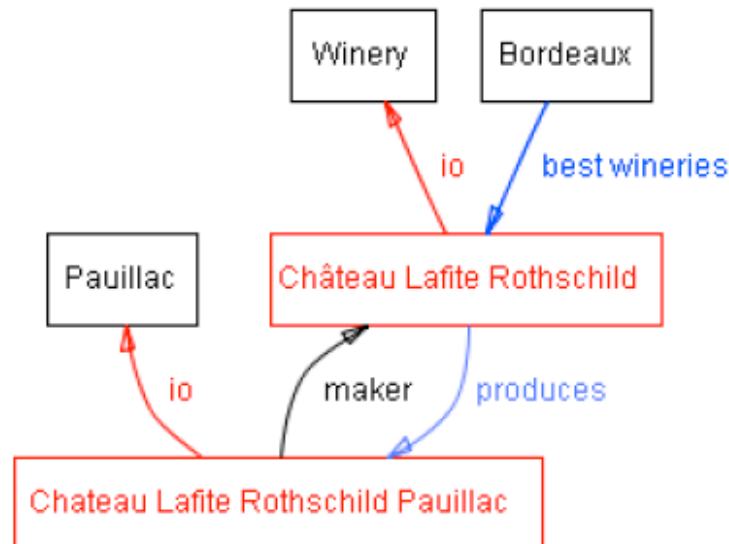


Figure 2.7 : Extract from Wine Ontology (Noy and McGuinness, 2001)

The semantic web is the World Wide Web Consortium's (W3C) vision of linked data that will support the development of machine readable data stores on the web that can be created using vocabularies, defined with data handling rules and accessed using

queries (W3C, 2015). The W3C does not make clear distinctions between vocabularies and ontologies, with ontologies commonly used to refer to any complex and formal structures of concepts used in the semantic web to assist in information integration. W3C technologies used for the development of ontologies include Resource Description Frameworks (RDF), RDF Schema and OWL (Web Ontology Language).

RDF provides a method of applying semantics to machine-readable documents without making any assumptions about the document structure (Fensel, 2003). RDF is intended to be used to express information about resources where a resource is any type of thing such as a person, location, document or abstract concepts (W3C, 2014). RDF statements about resources are made using triples that comprise of subject, predicate and object. An International Resource Identifier (IRI) is a machine-readable identifier of a resource that can be used as subject, object or predicate, for example

- The IRI “http://dbpedia.org/resource/Leonardo_da_Vinci” is an identifier for the person Leonardo Da Vinci
- The IRI “http://xmlns.com/foaf/spec/#term_knows” is an identifier for a relationship of personal acquaintance between two persons who know each other.

A literal is a resource that is not an IRI but has a data type such as strings, data and number which enables them to be parsed correctly. Literals are used only as objects in a triple. A resource may be referenced in multiple triples which results in a directed graph made of RDF triples that can be queried using SPARQL (SPARQL Protocol and RDF Query Language). The RDF Schema language allows the definition of semantic characteristics of RDF triples in terms of a vocabulary. An RDF Schema provides the components required to build an ontology such as class, properties, sub-class, restriction, domain and range.

Figure 2.8 provides example of RDF triples and how they combine to form a graph that relates concepts to each other.

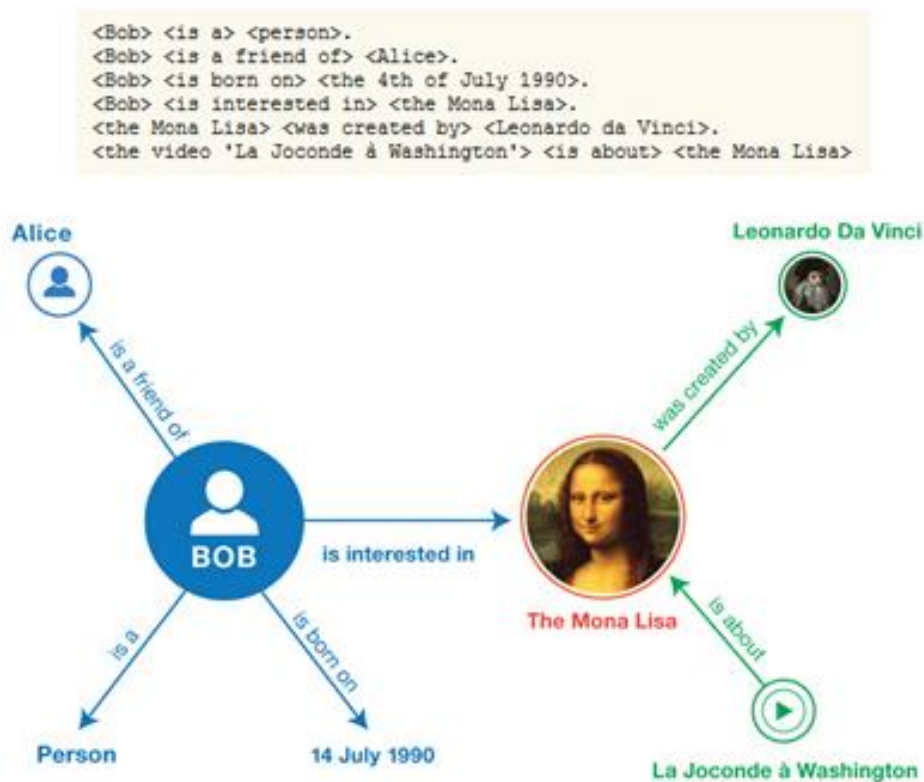


Figure 2.8 : RDF Triple pseudo-code and Graph Visualization (W3C, 2014)

OWL is an ontology language for the semantic web that be used along with RDF data and are primarily shared using RDF documents (W3C, 2012). OWL 2 is the current version developed by the W3C and is designed to support ontology development with the goal of making Web content more accessible to machines. At the core of an OWL ontology is a formal structure expressed with UML and an RDF graph. Semantic meanings are assigned to OWL ontologies either through Direct Semantics or RDF Semantics. The diagram in Figure 2.9 illustrates how a core abstract OWL ontology is expressed in documents with concrete syntaxes such as OWL/XML, RDF/XML so that they can be exchanged between applications. Typically users of OWL work with one syntax and one semantic layer.

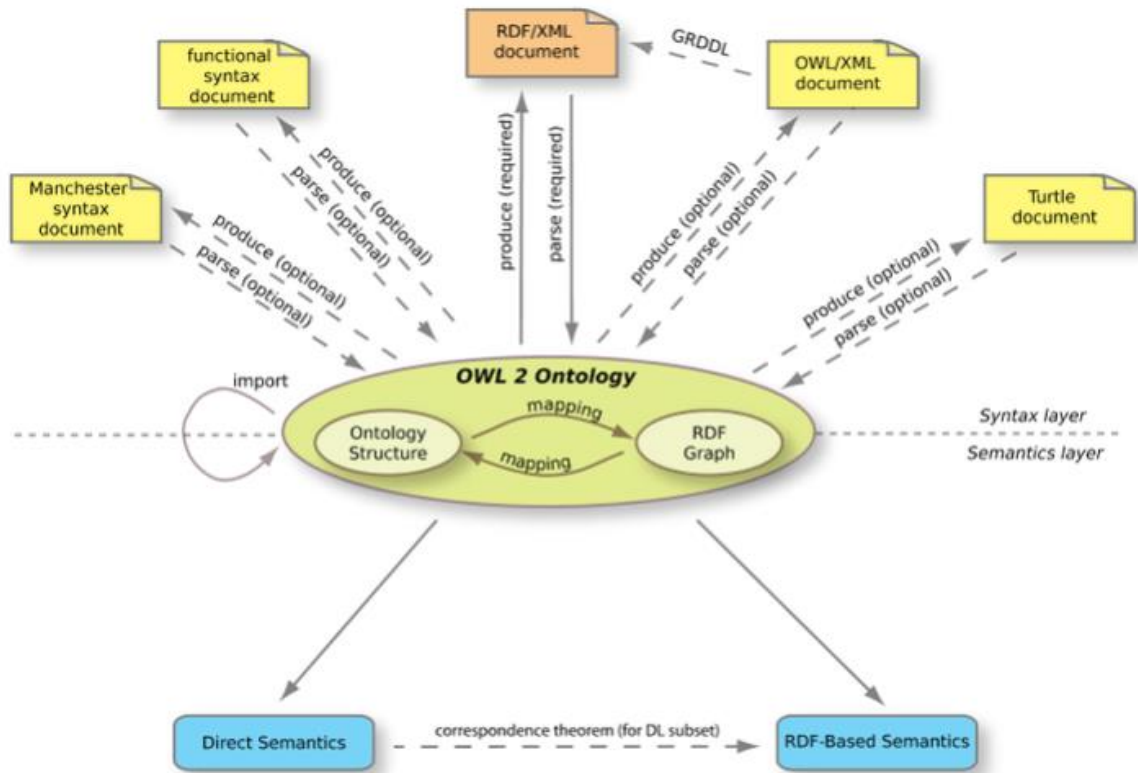


Figure 2.9 : Structure of OWL (W3C, 2012)

Smith *et al.* (2006) suggest that the use of technical and formal standards such as OWL do not guarantee the creation of an error for ontology that correctly represents the knowledge domain. An ontology is likely to be of little use unless it can engender its adoption across the domain so that its use becomes common practice. They find that development and adoption of ontologies in the area of biomedical informatics has encountered difficulties as it is not clear if purpose of the ontology is to be a controlled vocabulary, a conceptual representation or a knowledge model. Most importantly the ontology must be applicable in reality e.g. being of benefit to the patient by help the treatment of diseases (Smith *et al.*, 2006).

The Financial Industry Business Ontology (FIBO) is currently being developed as a repository of financial industry terms, definitions and relationships. It is being created by the non-profit financial industry trade association of the Enterprise Data Management (EDM) Council with technical guidance from the Object Model Group (OMG). The EDM Council (2014) states that the purpose of FIBO “is for data harmonization and for the unambiguous sharing of meaning across data repositories”.

The FIBO Semantics Repository¹ provides sections of the ontology that are being submitted to the OMG as part of the proposed standard. This repository is a canonical reference ontology optimised for semantic technology applications and is provided in RDF/OWL and UML formats. An extract from the FIBO repository is shown in Figure 2.10.

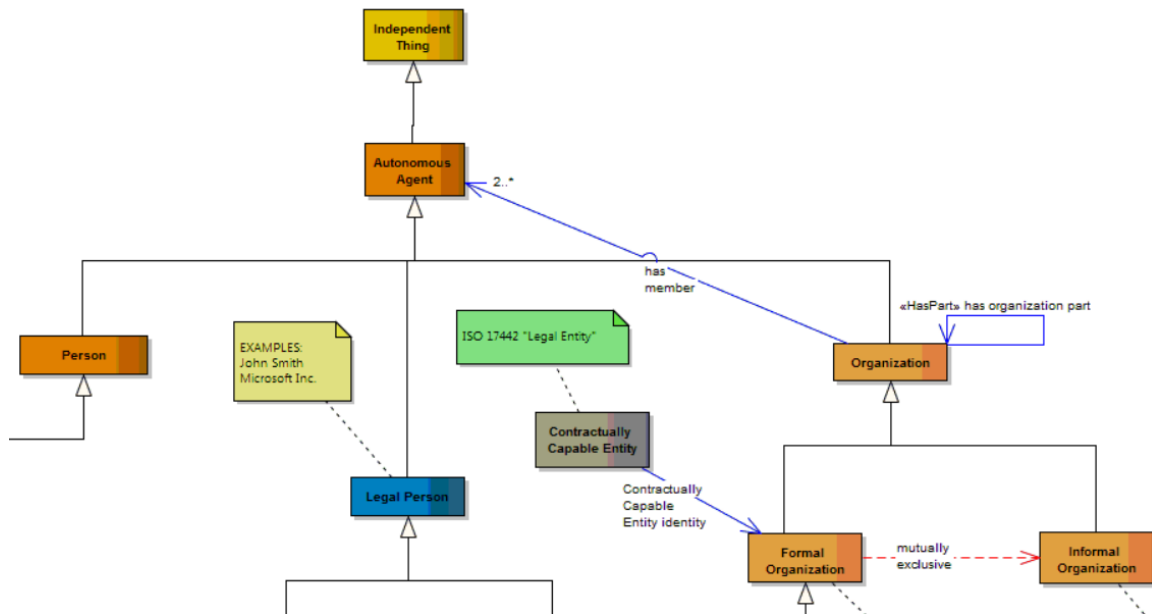


Figure 2.10 : Subset of FIBO Business Entity Diagram (EDM Council, 2014)

The developers of FIBO recognise the need to balance the requirements of an ontology to be conceptually formal and valid, operationally applicable in IT systems and yet grounded and accessible to users in the business domain. Mike Bennet (2014), Head of Semantics at the EDM Council, emphasises the importance of not attempting to describe the formal model structures to business users. A business view of the model should be consumable by the business user directly, without a modeller having to first describe the structure of the model first. Bennet (2014) uses the diagram shown in Figure 2.11 to emphasise that for an ontology model to be successful it must be effective in three dimensions

1. It must be based on a firm foundation of formal model structures
2. It must be implementable in information applications for the purposes of messaging and data interchange.
3. It must be grounded in the business domain that it is meant to support.

¹ <http://www.edmcouncil.org/semanticsrepository/index.html>

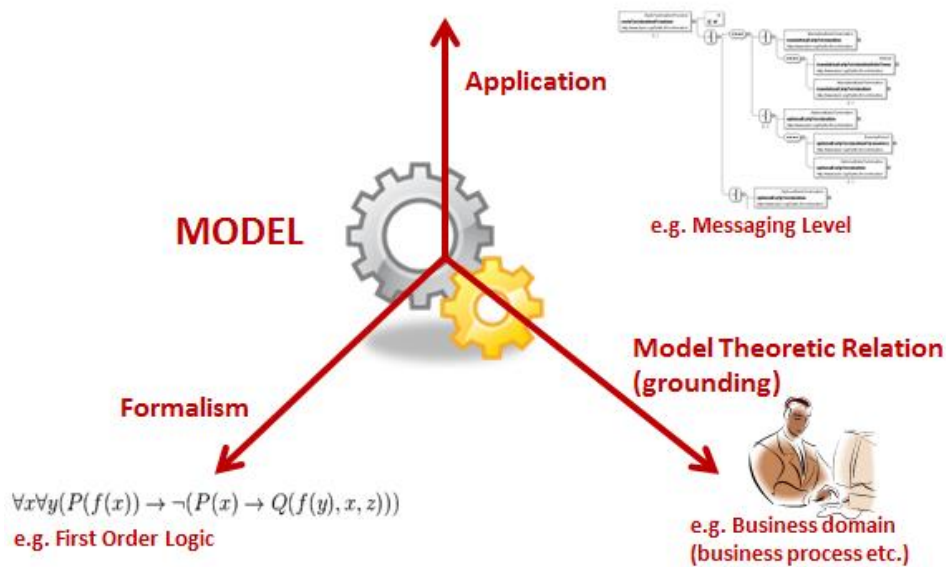


Figure 2.11 : Dimensions of an Ontology Model (Benne, 2014)

2.6 Comparison of Semantic Models

The description of a semantic model is often accompanied with a comparison of other types of the models from the semantic spectrum. This is either to illustrate the difference of the model by comparing their structure and purpose or to propose a method for translating one model type to another. Bennet (2104) suggests that an ontology can be validated through an iterative implementation in logical and physical models. This comparison also clearly puts ontologies and data models on different sides of a business and technology language interface. It can be implied that the ontology is analogous or a replacement of the conceptual data model of the Zachman framework.

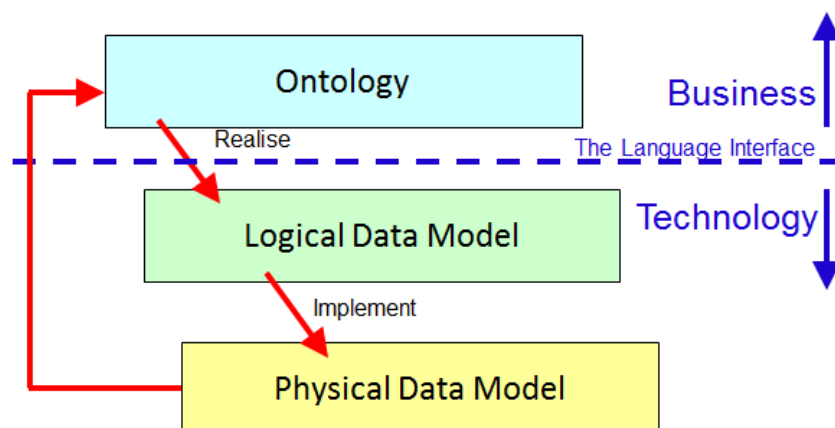


Figure 2.12 : Ontology as Conceptual Model for Data (Bennet, 2014)

Albarrak and Sibley (2009) propose an approach for generating an ontology for a business domain by deriving it from a relational model that already defines the main data elements and characteristics of the domain. They suggest that the ontology and the data model represent similar information in different structures. Their framework translates the data definition language (DDL) of the database into OWL that can then be refined by an ontology modeller. Their approach does not make use of descriptive comments for database objects and does not examine the data within the tables. The resulting ontology has a formal structure that matches the relational model but has no class instances inferred from the database. Albarrak and Sibley’s mapping of relational model to ontology is summarised in Table 2.2.

Table 2.2 : Mapping of Physical Data Model to Ontology Elements

Relational Physical Data Model	OWL Ontology
Table	Class Instance
Column	DatatypeProperty
Column Data Type	Range
Primary Key Constraint	InverseFunctionalProperty
Foreign Key Constraint	Object Property
Used to infer relational generalisations or sub-types	Is-a relationship, where the tables is a sub-type
Not Null	MinCardinality
Unique Constraint	InverseFunctionalProperty

Graudina and Grundspenkis (2011) suggest that OWL ontologies and C-Maps are both used to represent domain knowledge and have enough comparable elements that allows for an algorithm to perform automated transformations as shown Figure 2.13. They identify that a significant structural difference is that OWL ontology classes have attributes and data type properties which have no direct equivalent in C-Map concepts which do not have properties. Their proposed algorithm makes extensive use of the linking phrase between concepts to determine when a concept is to be converted into a ontology elements. They identify the importance analysing all of concepts ancestors if the semantic value of the C-Map is not to be lost when transforming to an ontology. The C-maps examined in their study are from an intelligent knowledge assessment

system that supports a limited number of linking phrases such as is-a, kind-of, part-of etc. These C-Maps are quite formal in their construction when compared to those proposed by educators such as Novak and Simon.

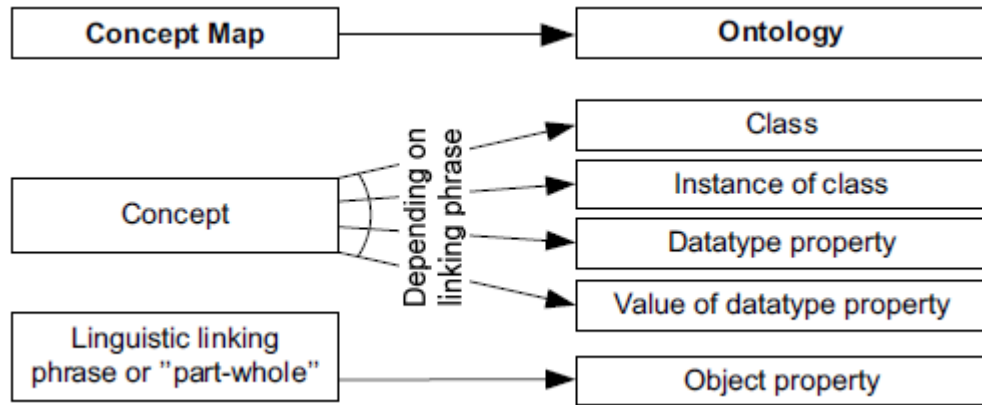


Figure 2.13 : Mapping C-map to Ontology (Graudina and Grundspenkis, 2011)

Obrst (2008) contrasts natural language terms which refer to real world with concept models that use entities and relationship to represent knowledge, as shown in Figure 2.14. A thesaurus allows for semantically weak term relationships such as equivalency while an ontology has strong conceptual relationships such as sub-type, part of, reliability relations, axiomatic rules.

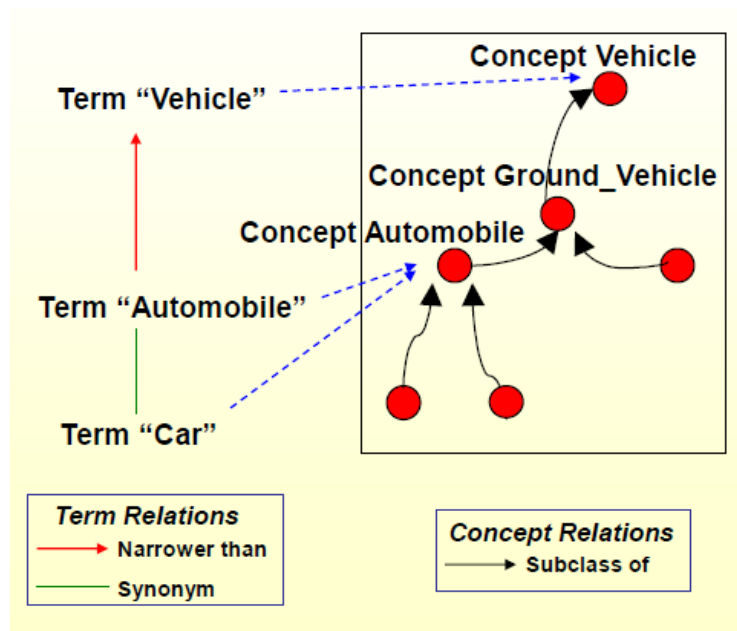


Figure 2.14 : Mapping of Terms and Concepts (Obrst, 2008)

Chisholm (2010) emphasises that a for a sematic modellers to correctly define a term they must understand that term definition will operate in the different contexts of the real world, the human mind and data management, as illustrated by the concept map in Figure 2.15.

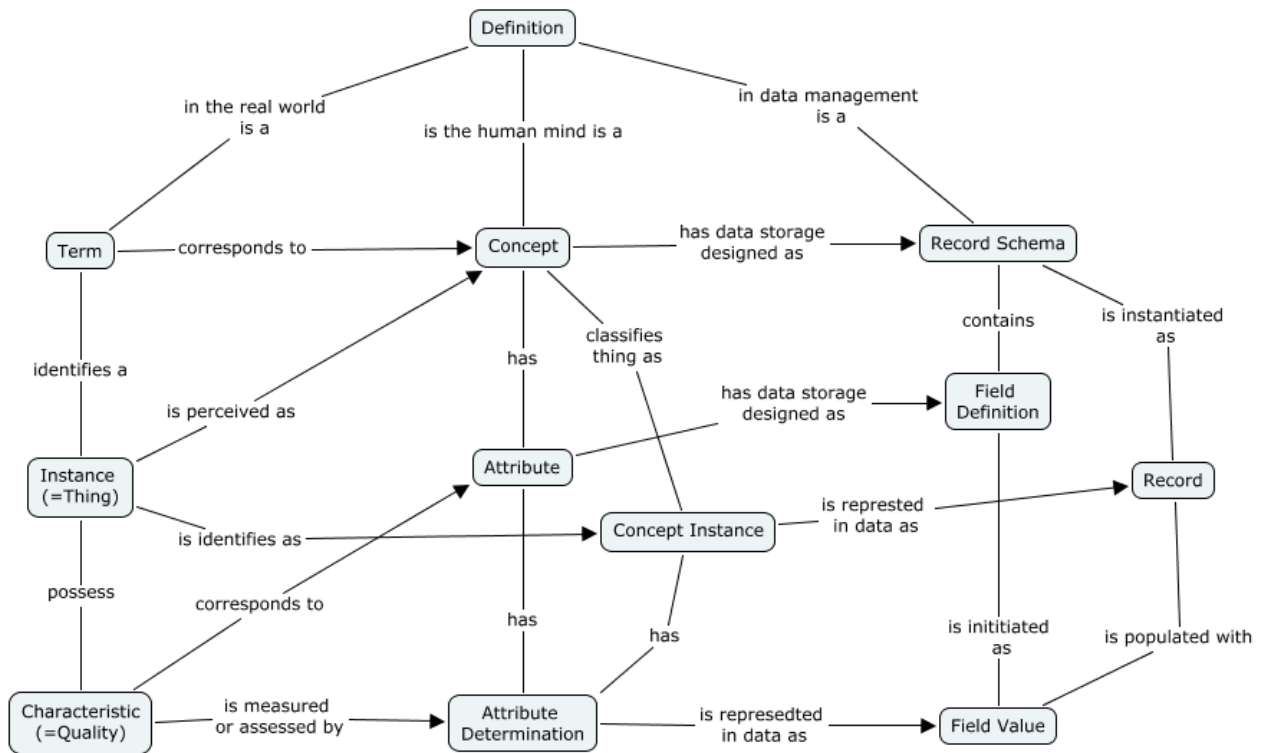


Figure 2.15 : Where Definitions are Found (Adapted from Chisholm 2010)

2.7 Conclusions

This chapter presented a review of the academic, technical and business literature that shows that there has been significant and active development of semantic models over the last 30 years. In that period that has been growth in the number of model types leading to somewhat confusing diversity of models which all share common goal of representing ‘things’ - objects, concepts, terms, entities – to assist communication of information between man and machine. The diversity has been driven by requirements for technical syntax for semantic models to match the evolution of information management systems, semantic web and knowledge management. However there is constant theme expressed that the different type of semantic models are related and are

facets of the same ultimate solution. The concept map in Figure 2.16 summaries the types and uses of semantic models discussed in this chapter.

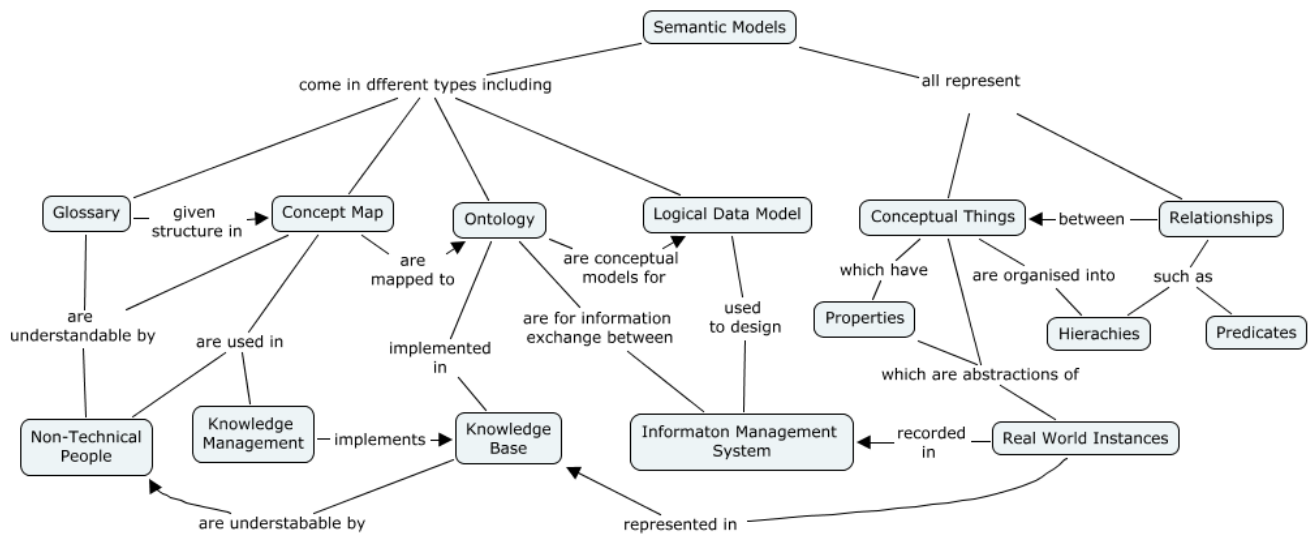


Figure 2.16 : Sematic Models Types in Chapter 2

The approaches proposed in the literature for the mapping of ontologies to both logical data models and concept maps support the approach of this research to implement a knowledge repository that combines these model types. The implementation of this knowledge repository is described in Chapter 5. However, the literature identified that a key factor for the successful implementation of the both ontologies and concept maps is that the modeller has an understanding of the knowledge they are representing. The literature review presented in Chapter 3 examines the relationship between knowledge management, knowledge acquisition and semantic modelling.

This literature review identified that ontologies are seen as the most flexible of these structures, possibly because they are designed to be readable by both machine and human readable. However the structure and formality of ontologies such as OWL can result in an artefact that is difficult for both technical and business users to understand. The developers of FIBO have identified the barrier that a user must to understand the structure of the semantic model before they can extract useful knowledge from it. This barrier can be overcome by ensuring that the end-users are exposed to more consumable versions of the semantic model. Chapter 6 provides an evaluation of a knowledge repository that combines the use of concept maps, definitions and ontological structures.

3 KNOWLEDGE MANAGEMENT AND SEMANTIC MODELLING

3.1 Introduction

This chapter provides an overview to the discipline of knowledge management and how it is related to the creation of the semantic models described in Chapter 2. Section 3.2 introduces knowledge management and makes comparison between the processes of knowledge modelling and semantic modelling. Section 3.3 describes the type of knowledge that is required to become an expert data modeller. Section 3.4 provides a summary of approaches and metrics used to evaluate data modeller knowledge.

3.2 Knowledge Management and Modelling

Knowledge management considers the interrelated themes of people, process and technology and how they relate to the implementation of knowledge management systems (Pope and Butler, 2012). Knowledge is commonly acquired by people either through communicating with experts who already hold the knowledge or by accessing written knowledge stored in a knowledge repository (Milton, 2007). Written or explicit knowledge is made available in a range of formats including books, electronic documents, websites, videos or specialist knowledge representation tools. Semantic models such as glossaries, concept maps and ontologies are example of tools used for the processes making knowledge explicit also referred to as knowledge modelling.

Knowledge management takes place in the context of human organisations that have procedures that support and constrain knowledge sharing between individuals. These processes create tacit and explicit knowledge through an iterative cycle of socialisation, externalisation, combination and internalisation that involves interactions between humans and knowledge (Nonaka *et al.*, 2000). This is similar to process of semantic modelling which involves a modeller eliciting domain knowledge from a domain expert for representation in an ontology or logical data model.

The most significant barrier to knowledge management is the changing of organisational culture to make the sharing of knowledge the norm (Blair, 2002). Such a culture is a prerequisite for effective communication between subject matter experts and knowledge modellers or business analysts. The acquisition of knowledge is

supported by a range of technologies including websites, blogs, document management systems and knowledge modelling tools. Knowledge management projects that primarily focus on technical implementations often fail as the technology becomes an end unto itself rather than a solution that successfully support knowledge sharing.

Knowledge modelling discussions frequently begin the question of ‘What is knowledge’ followed by a discussion that contrasts data, information and knowledge. Knowledge is described as either the conceptual understanding acquired by a person as they refine information (Rowley, 2007), or as the procedures that a person follows to correctly apply information in a given context (Gurteen, 1999). Milton (2007) suggests that knowledge can be categorised using the dimensions of explicit vs. tacit and conceptual vs. procedural. This approach is used in Table 3.1 to categorise examples of knowledge a person might have about bank customers. Each type of knowledge has associated types of representation or models. The types of semantic models discussed in Chapter 2 are associated with representing explicit conceptual knowledge.

Table 3.1 : Models or Representation of Types of Knowledge

	<u>Explicit Knowledge</u>	<u>Tacit Knowledge</u>
	Gained through reading or education and not difficult to write down and explain	Gained through practice or personal experience and difficult to explain and write down
<u>Conceptual Knowledge</u> Concepts and how they are related to each other	<u>Models:</u> Taxonomy, Concept Map, Ontology Example: ‘I know that the bank has customers and that each one has a credit rating’	<u>Models:</u> Beliefs & Biases, Mental Models Example: ‘I know that some customers in this branch manipulate their credit rating’
<u>Procedural Knowledge</u> How to perform steps to complete a task	<u>Models:</u> Process Map, Algorithm, Demonstration Video Example: ‘I know to calculate the credit rating for a customer’	<u>Models:</u> Heuristic, Intuition Example: ‘The way this customer is applying for their loan makes me think they are up to something’

The process of both semantic and knowledge modelling involves humans interacting with other humans in order to produce an explicit representation of knowledge. Knowledge modelling is part of the iterative knowledge creation processes described by Nonaka *et al.* (2000) as the SECI cycle, which is illustrated in Figure 3.1.

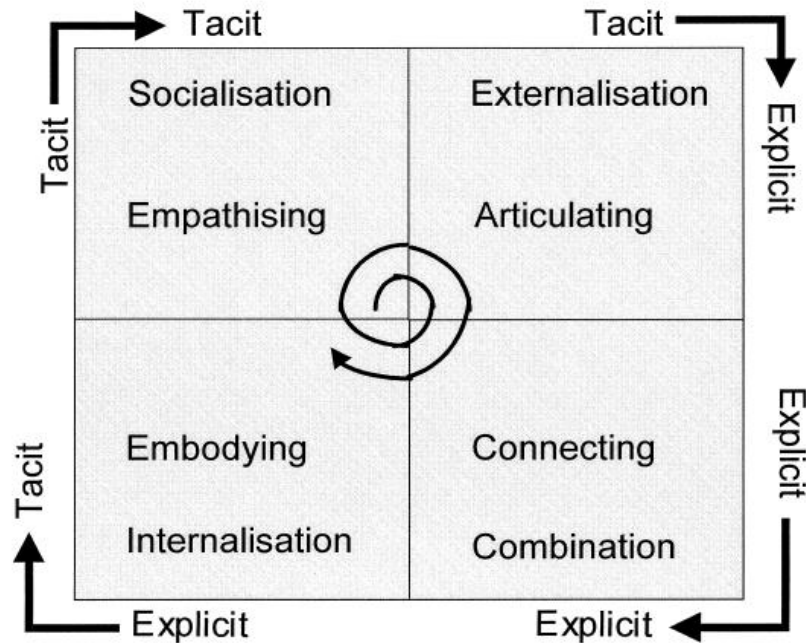


Figure 3.1 : The SECI Process (Nonaka *et al.*, 2000)

The four stages of the SECI process can be related to semantic modelling tasks as follows;

1. *Socialisation*: Tacit knowledge is shared between modellers and domain experts through natural techniques such as conversations, interviews and workshops. New tacit knowledge is created as experiences are shared.
2. *Externalisation*: Modellers articulate their tacit knowledge as an explicit model using a specific modelling technique. The new explicit knowledge is shared and requested from other modellers and domain experts for comment and feedback.
3. *Combination*: The modeller combines the new semantic model with previously developed complex models. This identifies concepts that are common between the models and also identifies new relationships. This may result in re-appraisal and rework as the modeller follows the structural rules required of the modelling technique.

4. *Internalisation*: The act of modelling is ‘learning by doing’ for both the modeller and domain expert. New tacit knowledge of the domain being modelled is acquired by all participants in the modelling process. The new explicit model is read by others so that they can use it to acquire new knowledge. This results in further socialisation of the concepts which starts the cycle again.

Correndo and Alani (2007) describe how an ontology based knowledge repository can be created by communities of knowledge workers using semantic web technologies and collaborative knowledge construction techniques. For large groups of users who are not co-located it is a requirement that the ontology tools support both remote and continuous participation by all members of the group. The required features of the tools for knowledge repository creation identified by Correndo and Alani can be mapped to the SECI process as follows;

- *Socialisation*: Support for discussion and consensus building between users such as instant messaging, discussion with annotations and voting or rating of concepts.
- *Externalisation*: Distributed editing functionality taxonomy editor that is available in real time to other users, common space for modelling
- *Combination*: Searching for existing terms to maximising reuse, tagging to help reuse, consistency checks where merging ontologies
- *Internalisation*: Visualisation tools to browse the ontology, Notes to provide advice on use of concepts

3.3 Knowledge Acquisition by Modellers

Venable (1996) describes how novice data modellers to become expert data modellers through the acquisition of specific knowledge and skills. He describes expert data modellers have extensive experience and/or training in both data modelling and a number of different business application domains. They possess a library of generalised data models contracts which they can adapt and apply to meet a variety information requirement. Their understanding of abstract data modelling concepts is acquired through technical training in different data modelling techniques such as entity relational modelling, UML or object modelling. In contrast, novice data modellers are described as having some formal training but very limited practical experience or

application domain knowledge. Typically a novice data modeller will have received education on entity relation modelling, system architecture and relational data modelling design as part of a computer science undergraduate degree.

Venable (1996) also contrast the modelling approaches of novice and expert modellers. A novice's approach modelling focuses on the minutiae of a data model problem in isolation rather the relating the problem to other existing parts of a data model. Experts take a more holistic view to ensnare that the data model is a coherent whole that supports the information needs of the solution. This viewpoint is supported by Moody and Shanks (2003) who identify that the early holistic analysis of a new requirement can help identify when then requirement can be met directly or by extending the existing data model. The most influential factor on data model quality is how understandable it is to by business users (Moody and Shanks, 2003). Expert modellers typically create more complex models then novice modellers and this is because the capture more of the requirements. There is strong negative correlation between simplicity and completeness (Moody and Shanks, 2003).

Venable (1996) also identifies that novices lack the high-level heuristics of data modelling that encourage a constant evaluation of the model such as 'look for ways to generalise existing structures rather than adding new' or 'stand back and reflect on the big picture'. Data modelling is an analysis and design technique that can be seen as an art rather than a precise science (Moody and Shanks, 2003). Venable suggests that the novice modeller's lack of these heuristics means that they do not have an awareness of the benefits of seeking feedback on the model and are more likely to compound one mistake with another. The novice data modeller is confronted with a hard task that requires them to apply newly learned data modelling constructs to solve an information requirement for a business domain they do not understand. This will be an uncomfortable experience for the novice as they may feel that they will appear stupid for asking obvious questions from colleagues whom they may not know very well.

Venable (1996) concludes that combination of these challenges results in a barrier that stops novice modellers becoming expert . Novice rush to complete modelling tasks and exhibit avoidance behaviours in relation to seeking feedback on their work. This barrier is an example of the knowledge acquisition bottleneck that is a common challenge to knowledge management projects (Wagner, 2006). The characteristics challenges of the bottleneck can be applied to the novice semantic modeller as follows;

- *Getting Started:* The modeller has to ask basic questions to obtain knowledge from business or end-user experts to understand the information requirement
- *Narrow Bandwidth:* The novice modeller lacks the heuristics to prioritise and generalise significant volumes of business requirements to turn them into information requirements
- *Latency of Acquisition:* The novice data modeller is being forced to acquire tacit knowledge of both the business domain and data modelling constructs. Time is required for them to internalise this new knowledge before they can make them explicit in the data model.
- *Inaccuracies in Knowledge:* Novice modellers tend to focus on the specific technical data modelling problem rather spending time to verify the domain meaning and consistency of the whole model
- *Maintenance Trap:* Modellers lack of experience have the time to maintain the model though feedback from end-users or other data modellers

Venable (1996) suggests that teaching strategies should be adopted for novice modellers so that the teaching of data modelling constructs is complimented with practice applying the constructs to multiple applications domains. Novice modellers should be encouraged to adopt expert modeller characteristics such as reviewing and critiquing the data models created.

3.4 Measures for Knowledge Acquisition and Modelling

An assessment of knowledge acquisition of modellers can be undertaken through a combination of directly assessing the tacit knowledge of the data modellers, measuring the quality of the explicit knowledge represented in the semantic model and evaluating the usefulness of the semantic model to end users. Milton (2007) suggests that the implementation of any knowledge base is assessed in terms of its impact and usefulness as an end-product. End user feedback can be obtained through a combination of questionnaires and interviews using to capture structured metrics as well as feelings and opinions. Where a knowledge base is implemented in a web site or application, usage statistics can be gathered on popular search terms, periods of user activity or frequently accessed knowledge objects.

Vieritz *et al.* (2013) implemented a knowledge map tool to assist computer science students in learning the concepts of object orientated programming and software

engineering over an eleven week period. The knowledge base provided explanation of concepts and terms illustrated with examples and visualised in graph that provided relationships between knowledge objects. The knowledge base was implemented in a custom ROLE e-learning environment and was accessed by students in the weeks coming up to the end of course exam. A screen shot of the knowledge base is provided in Figure 3.2.

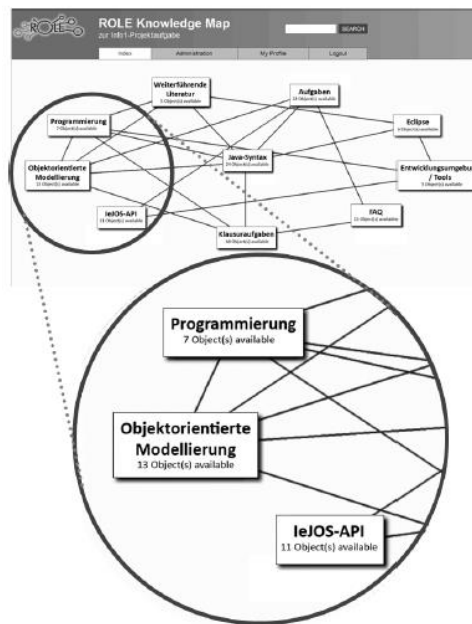


Figure 3.2 : Screenshot from E-Learning Environment (Vieritz *et al.*, 2013).

Vieritz *et al.* (2013) evaluated their knowledge base by asking both students and teaching staff to rate several statements on a scale from 1 (strongly disagree) to 5 (strongly agree) and explain their ratings. Students were asked about the usefulness of the e-learning environment and gave a mean response of 3.5 with standard deviation of 1.3. Teaching staff were asked to rate the statement ‘I would use such environments more often for learning if I had access to them’, with a mean result of 3.3 and standard deviation of 1.8. The evaluation also identified that some user would prefer printed material to online material as it better suits their learning style of using pen and paper to do exercises.

Simon (2007) used a combination of curriculum (high-level) and topic (low-level) concepts maps while teaching a course in financial accounting theories. The concepts maps were used during lectures and tutorials. Students were asked to complete an anonymous questionnaire that asked them respond to questions on a five point scale from 1 (strongly agree) to 5 (strongly disagree). It was observed that students found

concepts maps useful and assisted in their learning of the financial accounting concepts. Students found that the higher level maps assisted them identifying concepts that were important to the topic but were maybe omitted by lecturer due to time constraints. Both levels of maps were found to help students create links between concepts and detailed explicit knowledge in text books.

Osman *et al.* (2011) proposed that quality metrics used software engineering can be applied to ontology engineering.. There two disciplines share common lifecycle such as requirements specification, conceptualisation, formulation and implementation, and both require participation from users to ensure that out meeting their functional or knowledge need. A study was conducted to evaluate the effects of using an ontology during the teaching of data modelling techniques to novice data modellers. The study introduced the ontology at various stages of the teaching and evaluation to different student groups. As part of the end of course evaluation both groups of novice modellers who were provided with an ontology on data modelling concepts were asked complete ERD modelling task. Students were then asked to rate the level of confidence in their model and rate how useful the model was using a Likert rating scale from 1 (negative) to 10 (positive). This was combined with the lecturer's evaluation of each student's knowledge of data modelling in terms of comprehension and completeness.

Osman *et al.* (2011) found no evidence that providing the ontology during teaching improved student comprehension, but suspects that the study was biased by the presence of a number of very strong students in the group that was not provided with the ontology. When ontology was made available during the assessment a significant improvement in both comprehension and completeness was observed. However, only 29% of students rated the ontology as useful (a rating of 5 or more out of 10).

The data model quality factors proposed by Moody and Shanks (2003) can be interpreted from a KM perspective as the success to which the explicit knowledge in the model correctly represents and combines the tacit and explicit knowledge acquired during requirement and data analysis. They found data model quality is best measured using a small number of qualitative measures combined with soft text feedback from model stakeholders. Basic quantitative metrics such as number of entities and relationships reuse percentage, development cost estimates can be useful for heuristics

that can be applied during data model reviews, but attempts to quantify data model quality can be counterproductive.

The majority of flaws in a data models can be traced to data modellers not correctly eliciting requirements and knowledge from stakeholders during the modelling process (Moody and Shanks, 2003). This is illustrated though the actions taken and outcomes observed when addressing typical data modelling issues;

- Requirements are missed because they are not understood or captured correctly by the data modeller. This is addressed by iterative model reviews to ensure that the tacit requirement of the business user have been correctly externalised by the data modeller.
- Unnecessary requirements introduced by the data modeller that was not asked for by the business user. This was due to data modeller's lack of business domain knowledge that led them to make assumptions about what user wanted.
- Duplication of data elements to support the same business information requirement. This is addressed by performing pre-review reviews before any modelling work.
- Data modeller develops correct or technically perfect models that are viewed as incomplete or inflexible enough by business users, or not implementable by the DBA.

3.5 Conclusions

The literature review in this chapter has shown that there that there is an overlap between semantic models and knowledge management, for example Milton's (2007) proposed use of the semantic models as part of knowledge management projects, or the comparison between database software engineering and knowledge engineering made by Osman et al (2001). There are also similarities between the challenges faced by novice data modellers and the knowledge acquisition bottleneck. The task of creating machine readable sematic models is a very human activity. It requires collaboration between business and technical specialists to make explicit knowledge that is both hard express and difficult to understand. Moody and Shanks (2003) suggestion that high quality data models can only be created by performing iterative reviews with stakeholders can be interpreted an application of Nonaka's SECI cycle. The methods

by which experience financial data modellers acquired their knowledge is one of the topics in the interviews summarized in Chapter 4.

Given that concepts maps have been shown to be beneficial in the teaching of business and technical concepts they should be a useful knowledge acquisition tool for data modellers. Modellers could also develop concept maps during model development and use them to verifying shared understanding with business users. Concept maps could also be used to document meta-models that could allow modellers to better understand how a business concepts maps to data model entity or ontology objects. The process and benefits of developing concept maps in parallel with ontology and data models is discussed in Chapter 5 as one of the lessons learnt from the knowledge repository implementation.

The complexity of semantic models and business domains means that it is unlikely that any one single modeller can have all the knowledge they require to perform modelling tasks stored as tacit knowledge in their head. Instead they need to be able to access knowledge by talking to colleagues and accessing semantic models based knowledge repositories. The literature suggests that the evaluation of both semantic models and knowledge repositories is best undertaken with a combination of qualitative and simple quantitative measures. The design and use of both these types of measures is described in Chapter 6.

4 FINANCIAL INFORMATION MODELLER INTERVIEWS

4.1 Introduction

This chapter provides a summary and analyses of interviews that were conducted with two financial data and semantic modelling subject matter experts (SME). The interviews were undertaken to compliment the background research conducted in Chapter 2 and Chapter 3 with the SME's opinions on semantic models in the financial industry and how data modellers acquire business domain and data model knowledge. Summaries of the interviews are provided in the Sections 4.3 and 4.4. The transcript of the interview was analysed using formal method and the results of this analysis are presented in Section 4.5. Section 4.6 describes the results of a short contrived knowledge elicitation was conducted with each of the interviewees.

4.2 Interview Approach

The interviews were arranged with two individuals who work for a large multi-national IT corporation and have extensive experience in the development of information models for the financial sector.

- SME 1 is the Product Manager for a range of information and process models developed for the financial and insurance sectors. She has over 20 years of experience in modelling and business analysis with an emphasis on banking and financial markets.
- SME 2 is the Lead Architect for a range of information and process models in financial, insurance, health and telecommunications sectors. He has over 20 years of experience in modelling and information architecture in the financial industry.

Milton (2007) suggest that a semi-structured interview style is an appropriate technique for this type of research activity as it allows the interviewer to informally elicit knowledge on a number of topics while focusing the expert on the knowledge relevant to the research questions. The interview questions are provided in Appendix A. Each interview was followed by a short contrived knowledge elicitation comprising of a card sort and a triadic exercise. The concepts for the card sort includes the

semantic models types of covered in the literature review, the artefacts included in the experiment and the human and IT actors associated with the research question..

The first theme of the interview was to explore the acquisition of technical and business domain knowledge by data modellers that were discussed in Chapter 3. Venable (1996) describes how individuals must acquire a combination specific knowledge and skills to become expert data modellers. Data modellers having the business domain knowledge to create models that are understandable to business users is as the most influential factor on data model quality (Moody and Shanks, 2003). The interview included questions to explore how the SME acquired their data modelling and business domain knowledge. Broader questions on this topic were used to obtain the SME's perspective on how novice modellers could go about acquiring business domain knowledge and the role that semantic models play in knowledge acquisition.

The second theme of the interview was to refine understanding of the types of semantic models used in the financial sector that were identified in Chapter 2. Concept maps have been used to in the teaching of financial concepts appear to be as widely adopted as ontologies or structured languages. The discussion in the literature also emphasises the conceptual and technical mapping between the different model types of semantic model (Bennett, 2014), (Albarrak and Sibley, 2009). The interview includes question that ask the SMEs to identify trends and the state-of-the-art use of semantic models in the financial sector. Follow-up question examine their experience and opinions on the use of semantic models and related modelling techniques.

4.3 SME Interview 1 - Models Product Manager

This section provides a summary of the interview with SME Interview 1. Direct quotes from the interviews are identified by the use of text in italics. The full text transcript is provided as part of the supporting material. The interview was 37 minutes in duration and was directly followed by a triadic and card sort exercise.

4.3.1 Acquisition of data modelling and business domain knowledge

SME1's data modelling knowledge was primarily acquired through practical experience developing data warehouse solutions supplemented by reading texts such as Kimbal's Data Warehouse Toolkit. Her financial sector business domain knowledge was acquired through a combination of research and through talking to colleagues who had

expertise in specific business issues. This typically involves conversation with experts who are asked by the modeller to “*tell me from the beginning what does this all mean or how does this all work together*” and then the modeller has to piece the new knowledge with what you already know.

When discussing these two types of knowledge, SME1 associated business domain knowledge with the role of a business analyst whose job it is to understand the whole business problem and figure out any impacts on a data model. A person in the data modeller role must have the technical knowledge to actually make those changes to the data models. However it is not uncommon that these two roles may be carried out by the same person.

One of the knowledge acquisition techniques highlighted was that as a modeller you have to be “*relaxed enough in your own knowledge that you can ask the more basic questions*”. She described how she would understand a concept by first developing a concept of how she think a business process works and would then test her understanding with people who she trusts to have the correct knowledge.

4.3.2 How novice modellers acquire business domain knowledge

SME1 assumed that a graduate or novice data would have received formal training in data modelling as part of their education. Novice modellers may start with basic level of business knowledge gained as a banking consumer such as opening a bank account or paying interest on a loan. However there would be “*an expectation that they will spend a couple of years laying down the basic and the fundamentals of knowledge of the business*”. Expert business domain knowledge is acquired by understanding financial processes not only from the perspective of the customer but also from the perspective of the bank, such understating how a bank gets access to funds in order to lend them to customers.

There is an incentive for financial institution to accelerate business domain knowledge acquisition because if data modellers make a mistake the bank is potentially going to lose money. There is potential for using industry ontologies such as FIBO to assist a data modeller in acquiring business domain knowledge in combination with the individual’s existing knowledge. This assumes that the semantic model has already put together by someone who can explain who how the financial industry works. The

visual nature of semantic models should also make them suitable for of any modeller to read.

4.3.3 Use of semantic models and related modelling techniques

SME1 describes semantic models as something that references real world concepts, their relationships and the associated rules about information. They are used to represent business knowledge, to get a better understanding of the elements of an organisation and to understand the world in terms of the financial markets. They are also used in the implementation of business rules in technical system to provide the control of IT applications and therefore the control of the business. Semantic modelling is about making sure that a modeller captures the rules of what the information is and how the different information concepts relate to each other.

SME1 has worked with a variety of semantic models;

- IBM Financial Services Data Model (FSDM), a classification style semantic model that captures the relationships between different concepts, but not information rules such as cardinality
- Semantics of Business Vocabulary and Rules (SBVR), a structured business language maintained by the Object Model Group (OMG). This allows for the definition of business rules using a language that is easier for the business audience to use.
- Financial Industry Business Ontology (FIBO), a UML based business ontology that support information rules such as cardinality and multiple inheritance. This is characterised as a technical model that is trying to be a business model.

SME1 described data models as having a restricted number of ways of expressing business and information rules when contrasted with semantic models. She would expect that when a data modeller encounters a semantic model they would “*understand that this is a map of the world in a certain language*”. If the data modeller can understand a semantic model it then essentially becomes a translation exercise for them to understand the business rules. She suggested that a data modeller would not actually need to understand of the business if they had an algorithm to turn a semantic model into a data model.

4.3.4 Trends and state of the art use of semantic models in the financial industry

SME1 describes a cycle in the financial sector that starts with a great belief in the benefits of semantic models but then those responsible for implementation take too long to actually make the model useful. Such delays result in frustration leading to data warehouse implementations that do not use a semantic model and instead take a “*Nike attitude, let’s just do it*” or “*just slap on a few more columns onto these tables*”. This ultimately results in a mess as the data models implemented do not integrate or do not correctly abstract the business concepts. This leads to a call for the use of semantic models and the cycle begins again.

SME1’s opinion is that we are currently at the stage in the cycle that emphasises the importance of semantic models. This is illustrated by the current industry wide effort in the development of FIBO which is an information modelling response to financial markets that are so complex and where people can just make up new products and trade them. The business imperative is that institutions and individuals lost huge amounts of money in the financial crisis because they did not understand the underlying instruments they were trading. The objective of FIBO is to provide “*a structure to understand exactly what the different instruments are and according to where they are in that hierarchy of instruments, you know some are more smelly than others, some are more risky than others, there is risk attached to every single financial instruments*”. The ontology becomes the industry language of common understanding through which regulators can specify regulatory rules and reporting requirements.

4.4 SME Interview 2 – Models Architect

This section provides a summary of the interview with SME Interview 2. Direct quotes from the interviews are identified by the use of text in italics. The full text transcript is provided as part of the supporting material. The interview was 36 minutes in duration and was directly followed by a triadic and card sort exercise

4.4.1 Acquisition of data modelling and business domain knowledge

SME2’s data modelling knowledge was initially acquired through theoretical training courses which he then put into practice implementing data warehouses. In-depth knowledge of the banking industry came from a combination of formal training in combined with many years hands-on experience working in the sector. Learning

business domain knowledge about banking and financial markets is challenging because “*you are never sure you know the full range because it's quite a large industry*”. His expert knowledge was acquired through a combination of personal learning by doing combined with working with other modellers – “*seeing what people are doing, understanding what they are doing and how that applies to what you are doing*”. This learning style is consistent with the SECI models of knowledge acquisition.

A clear distinction is made between the knowledge required by individuals who create data models and those who use or read data models. A data modeller is characterised as “*someone who is three days out of five is doing something on data models*” using one of a variety of data modelling tools such as Information Data Architect, ERWin or Power Designer. Business analysts are “*slightly more looking at the requirements that feed into the model*”, while Information Architects look “*at the broader technical ecosystem that once again influences or underpins the data model*”.

4.4.2 How novice modellers acquire business domain knowledge

The most effective way to teach business concepts is not to directly use a model. It is better to use informal approach such as a “*very large white board and you start explaining the concepts as a set of bubbles*”. A semantic model can be the basis for teaching business knowledge but rather than just providing the model to beginners without any explanation, it would be more advisable to use the model as the language to teach people the business concepts. A simple business example should be used as the hook for using the language of the semantic model to describe a concept familiar to the novice. This example is then used to explore orthogonal concepts from both the banking consumer and the financial organisations point of view.

Reading an ontology could be a way to gain business domain knowledge, but the novice would have to start with a frame of reference or a hook to get into the ontology. For a complex area like Basel that assumes quite a high degree of business knowledge, a novice modeller would first have to build up their own knowledge by reading and researching documentation. The semantic model would help to extend their knowledge of the specific details or classification constructs.

4.4.3 Use of semantic models and related modelling techniques

SME2 describes semantic models as a group of business user friendly models that define concepts and relationships. They are less technical than entity relationship data models or UML models. They are “*aimed more towards either communicating to the business or trying to lay down what it is that the business is trying to say*”. The primary use of semantic models is in defining information architectures and information governance processes. A semantic model should not be developed for individual areas of a business but instead it should become the semantic model for the complete enterprise.

However there is no clear accepted definition of semantic model with the term being used to describe ontologies, hierarchical models, vocabularies and glossaries and if “*you asked 10 different people you are going to get 15 different answers*”. The lack of agreed definition leads to difficulties when integrating semantic models with other types of models into architecture frameworks such as The Open Group Architecture Framework (TOGAF).

One of the challenges of developing ontologies is “*getting the balance right between machine readable, and I don’t meant you [a data modeller], I mean business person human readable*”. The difficulty with ontologies is that “*there is kind of an identity crises going on in that people look at ontologies and think it is machine readable and is something from which you can drive a business*”. His experience is that hierarchical or taxonomy models whose primary purpose is to be human readable are more likely to be adopted by business users.

The boundary between semantic models and regulatory reporting is a good example. A semantic mode can describe both the Basel regulatory reports and the information patterns that support and connect them. There is value of building or buying a set of semantic models that capture the inherent information and relationships patterns out of regulatory documentation.

4.4.4 Trends and state of the art use of semantic models in the financial industry

There is a trend in recent years where financial institutions expect a tighter integration between business vocabulary or semantic model with the underlying technical models. There is a growing appreciation of the role of a semantic model as a tool to assist in the governance across an information landscape that is far more complex than it was 10

years ago. This expectation is driven by both regulatory pressures on organisation to be able to demonstrate that they understand both their operational data lineage, and the need for efficient development processes that transform information specification into implemented IT assets.

There is a cultural change away from the days when banks could afford, or be seen to afford expensive data architecture organisations who would spend lots of time working on these wonderful precise models that end up taking six or nine months to implement. There is a “*very strong focus on people needing to be seen to, as the American's would say, drive out business value or time to value, to get things up and running quicker, faster, cheaper*”. Industry wide collaborations such as FIBO could help organisation meet both the expectations if they are successful in what they are hoping to achieve. The lure of such industry wide ontologies is that they correctly capture the business concepts and rules, while also being ready for immediate deployment in an “*automated or at least semi-automated way very quickly into the rules driven applications in different technologies*”.

4.5 Analysis of Interviews

While the summary and the text provided personal viewpoints of interest to the research question, a formal knowledge analysis was conducted that identify concepts and themes that could be used in an evaluation with the finding of the literature reviews in Chapters 2 and 3. The text of the interviews was analysed using structured qualitative methods to identify the most important semantic modelling concepts mentioned by the subject matter experts. Transcripts of the interviews were transcribed from audio to a text documents and then imported into the text analysis tool MAXQDA. The transcripts were analysed and a coding scheme was created to summarise the knowledge elicited from the SMEs.

The knowledge analysis resulted in a concept coding scheme that grouped the concepts discussed by the SMEs into four high level categories;

1. The actors or role involved in the creation and use of semantic models
2. Methods used by data modellers to acquire knowledge
3. Different types of semantic models
4. Challenges encountered implementing semantic models

An analysis of the concept codes, categories and count of the number of times each concept code was used in the transcript is provided in Appendix B. The coded transcripts are available as part of the sporting material. The following sub-sections describe the concepts identified in the analysis.

4.5.1 Actors in Semantic Modelling

The SME's categorised the actors involved in the development and use of semantic models as into business users, modellers and IT Systems.

- Business, including line of business users, regulators and industry bodies. There are the ultimate consumers of the information that is being generated by IT systems. There is also an expectation that they will be able to use the concepts and language of a semantic model.
- Modellers, including data modellers, business analysts and information architects who develop the range of semantic models. A single individual will frequently play multiple roles at the same time.
- IT System, including the application, databases and tooling that supports information and semantic modelling. There is a strong emphasis on the influential role played by modelling tools such as Protégé, Ab Intito, Power Designer and Information Governance Catalogue.

4.5.2 Data Modeller Knowledge Acquisition

The SME's view is that data modellers acquire knowledge through both formal and experienced based methods.

- Formal Learning such as class room training that is used to acquire basic knowledge of data modelling techniques and industry specific concepts.
- Informal Learning used to obtain specialist knowledge using methods such as reading, learning-by-doing and conversations with data modellers and business users.
- Industry Modelling Expertise which is based on many years of data modelling experience and is expressed by comparing how a novice and expert modeller would deal with different situations. This is a combination of both business and technical knowledge based on a broad knowledge of the industry matched with specific semantic model implementation experience.

4.5.3 Types of Semantic Models

The SME's referred to semantic models using both generic and specific terminology.

- A semantic model is generically described as something that captures rules for structuring, communicating and governing business information or concepts. There is a wide verity of semantic model types and it can be difficult to get modellers to agree what exactly a semantic is.
- Ontologies are formal representations of information that is both humans and machine readable. Their strong modelling structures such as sub-types, relationships and properties are currently used in the financial sector to create FIBO. Ontologies are associated with the OWL and RDF formats, and tools such as Protégé.
- Conceptual models are less formal then ontologies and would include structured business languages and classification models. They are primarily used with business users to capture and communicate information requirements. Examples include SBRV and other common business languages.
- Data models are used to define the structure of relational databases. While logical data models share some of the formal structures of an ontology, they do have all the structures required to specify complex business rules.

4.5.4 Challenges in Adopting Semantic Models

The SMEs described the challenges that they have experienced or observed in the adoption of the semantic models by the financial sector.

- Limited acceptance of semantic models occurs with both technical and the business user groups. Acceptance by business users can be improved when the focus is on a common business language that has a hierarchical structure and is made available in visual formats in a web browser, as such a model is not overly complex in its presentation or tooling. The challenge of acceptance with technical communities is that semantic models such as ontologies do not neatly fit into existing architectural frameworks which leading to confusion as to where and how the work with existing models.
- Semantic models are hard to implement. While there are technical standards such as RDF for delivering ontologies there is no standard method or set of tools for expressing or implementing the models. Getting technical users to

accept and use the models can be aided by creating easy to understand example that teach them the formal language of the model and then ask them to use the model to perform a tasks.

- Difficulties representing business knowledge in a semantic model. The more complex and refined the semantic model becomes the more it actually does capture the reality of how things are different. There is a skill in creating a single model that can be both consumed by a general business audience but also represents the detailed knowledge required by specialists such as brokers. This difficulty is compounded when an ontology is required to grow in order to represent the knowledge from the all business areas typically found in large financial organisation.
- Semantic models bridge the business and IT gap and frequently get caught up in political turf wars for which part of the organisation owns the definition of business and information rules. There are often conflicting business drivers for implementation of semantic models as they expected to both improve operational time-to-value and improve information governance. On the one hand there is a very strong focus on the need drive out business value by getting new business solutions up and running quicker and cheaper at an individual project level. At the same time larger organisations is trying to understand that complex the information landscape into which all the smaller solutions are being deployed into and they need to have a model to ensure that everything is integrated correctly.

4.6 Analysis of Contrived Knowledge Elicitation

Each interview was followed by a short contrived knowledge elicitation comprising of a triadic and a card sort exercise. This approach was used to elicit additional knowledge from the SMEs that could help refine the design of the experiment described in Chapter 6. The concepts for the card sort include the semantic models types of covered in the literature review, the artefacts included in the experiment and the human and IT actors associated with the research question. Two card sorts were conducted after the triadic to elicit further categorisation of the selected twelve concepts. The detailed results of the triadic and card sorts are detailed in Appendix C.

The results of the elicitation emphasised the importance of the role played by the individuals involved in the development of semantic models. Data Modellers are designers who use semantic models such as ontologies in the development of operational IT artefacts such as relational models, glossaries and reports. Business users utilise semantic models to help them express business and regulatory knowledge in a way that is understandable to both technical and non-technical individuals.

The concept map in Figure 4.1 is representation of the SME’s option of semantic models based on the results of the knowledge elicitation. It shows that semantic models are created by data modellers to bridge the divide between the business and technology users. The emphasis in this concept map on the needs of the individuals associated with the different types of semantic model. This can be contrasted with the semantic model concept map in Figure 2.16 of Section 2.7 which emphasises the implementation of semantic models to classify and manage information.

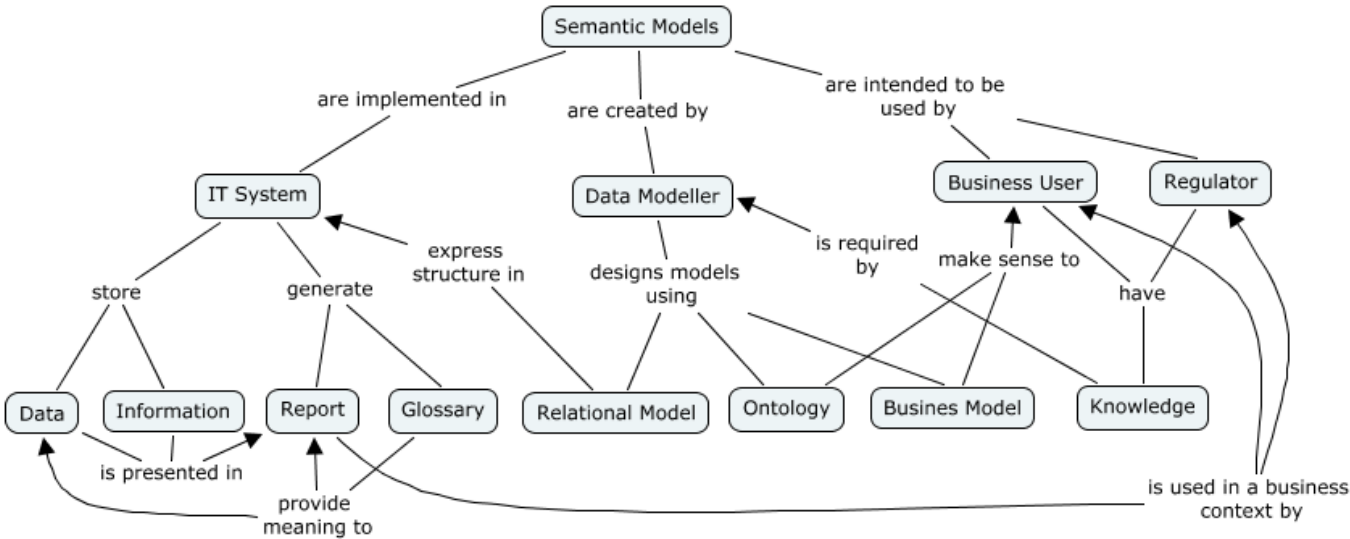


Figure 4.1 : Concept Model based on Contrived Knowledge Elicitation

The contrast reflects the SMEs experience of being data modellers who are expected to understand business domain knowledge that is typically in the heads of business users. This creates chicken-and-egg situation when creating semantic models as the data modeller is expected to create semantic models that communicate business knowledge to business users before the modeller understand the business concepts themselves. At the same time the models are expected to the be suitable to be used as the basis for implementing the technical solutions that are used for business operations.

4.7 Evaluation of Interviews

This evaluation discusses the use of semantic models in the financial industry by contrasting the experiences of the subject matter experts (SME) with the findings of the literature review in Chapter 2 and 3. The following sub-sections address research questions 1 and 2 that were introduced in Section 1.5.

4.7.1 What types of semantic models are used in the financial industry?

SME1 was of the opinion that it can be problematic to get consensus of what exactly is a semantic is or what is not - “*you asked 10 different people you are going to get 15 different answers*”. The SME’s identified that the semantic models used in the financial industry fall into the two broad categories of formal models and conceptual models. Formal models such as ontologies are a representation of information that is both humans and machine readable. Such strong modelling structures are being used to create FIBO and are associated with the OWL and RDF formats, and tools such as Protégé. Less structured or conceptual models include structured business languages such as SBRV or classification models such as IBM FSDM. These informal models are primarily used with business users to capture and communicate information requirements. Their ranking of semantic model by their ability to formally express machine implementable business rules echoes both the spectrum of models described by Obrst (2003) and the review of models types by Kramar (2013).

The SME’s view on the correct approach to successful use of semantic model is similar to that of Mike Bennett, Head of Semantics at the EDM Council, who empathises the importance of semantic models as a communication tool that must be meaningful to both business and modellers (Bennett, 2014). The combination of a structured business language such as SBVR and ontologies reflects the discussion of the Abi-Lahoud *et al.* (2014) on the development of complementary semantic models for financial sector. However the SMEs were not aware of concepts maps as a type of semantic model, which comparable with the literature review in Chapter 3 which found that concepts maps are only discussed in relation to financial education such as the studies described by Greenberg and Wilner (2015) and Simon (2007).

Research Question 1: What types of semantic models are used in the financial industry?

Evaluation: A wide variety of semantic models types have been applied in the financial industry over the last twenty year driven by changes in regulatory reporting requirements, semantic modelling formats and numerous financial crises. Ontology based initiatives such as FIBO attempt to become the language of common understanding between financial institution using formats that are both machine and human readable. Structured vocabularies such as SBVR are used to improve communication with business subject experts and technical experts. Ontologies and structured languages can be used in a complementary fashion and it is possible to translate the content of one format to another. Concept maps are not widely used in the financial industry but are used in some areas of financial education.

4.7.2 Are semantic models used by data modellers to acquire knowledge?

The SMEs agreed that both business and technical knowledge were required to create data models, and that it was expected that a data modeller would have to acquire a reasonable level business knowledge after many years of experience. It is not uncommon for the one individual to act in the roles of both business analyst and data modeller. The expectation that an experienced modeller would typically have the knowledge to act as both a technical data modelling specialist and business analyst contrasts with the data modelling literature which predominantly emphasises the technical aspects of data modelling. Typically the role of the business user or analyst who provides information requirements is seen as very separate from the technical roles of data modeller, information architect or database administrator (Moody and Shanks, 2003) (Blaha, 2010).

The SME's knowledge of technical data modelling and the financial industry has been gathered through a combination of formal class room settings and informal learning from extensive individual research and conversation with business or technical subject experts. This description supports the analogy of Nonaka's SECI process described in Chapter 3. Data modelling knowledge is acquired through formal training and solution implementation experience, while business domain expertise is acquired through conversations that are integrated with mental models. SME1 identified that for a data modeller to acquire knowledge one has to be "*relaxed enough in your own knowledge*

that you can ask the more basic questions". Having confidence in knowledge to ask the simple questions is one of the expert modelling characteristics identified by Venable (1996).

Both the literature and interview agree that a common challenge when creating semantic models is that modellers do not have the required business domain knowledge. The contrived elicitation identified that an individual in the role of data modeller is often the person faced with the translation problem discussed by Abi-Lahoud *et al.* (2014). Neither of the SMEs have used or considered using a semantic models as a method of acquire business domain knowledge, but both saw possible benefits in taking this approach. SME1 suggested that a structured and visual knowledge representation in an ontology such as FIBO should be accessible to technical modellers, and that adopting this approach could help financial institution accelerate learning within their technical teams. SME2 agreed but warned that a novice user could struggle with a complex area like Basel regulation if they did not have at least a frame of reference with respect to internal bank operations or financial regulation. A contrasting approach was identified by SME1's who suggested that a data modeller may not actually need to understand the business domain if they had an algorithm to turn an existing semantic model into a data model. This is similar to the suggestion made by Abi-Lahoud *et al.* (2014) that a semantic technology expert could successfully create an ontology by transforming business knowledge that been encoded in Structure Business Vocabulary and Rules (SBVR) by business experts.

Research Question 2: Are semantic models currently used by data modellers to acquire knowledge?

Evaluation: There is an emphasis in the financial industry on the use of semantic models to ensure clear communication of information requirements between technical modellers and business users. Semantic models do not appear to be widely used by data modellers to acquire knowledge to assist them in the task of translation business requirements into technical models. This emphasis on the use of semantic models primality as an implementation tool rather than a knowledge repository is at variance with the expectation that data modellers must acquire business domain knowledge as part of developing the professional expertise.

4.8 Conclusion

This chapter presented a summary and analysis of interview conducted with two data modellers who each have more than 20 year of experience developing models for the financial sector. The interviews captured their views on the practices challenges and trends in the use of semantic models and data models. SME 1 highlighted the regulatory drivers that heavily influence the trends in the use of semantic models in the financial industry, while SME 2 emphasised the technical aspects of the semantic modelling and the challenges of fitting them into existing information architectures.

The SME's view of the range and applicability of semantic models in the financial industry agrees with the literature reviewed in Chapter 2. They emphasize that semantic models have to be evaluated in terms of how useful they are to business users, but did not suggest that a semantic model could be equally useful to data modeller in acquiring business knowledge. The challenges of using semantic models as a knowledge repository are reflected Chapter 5 which describes the implementation of complementary ontology and concept maps

Their description of how they acquired knowledge is comparable with the SECI process and characteristics of expert data modellers described in Chapter 3. While they did agree that semantic models could be used by used as a knowledge repository for data modellers, there was a warning that combination of a complex ontology and unfamiliar specialist knowledge could result in a knowledge repository that is inaccessible for novice financial data modellers. The experiment described in Chapter 6 explores this concern by involving both experienced and novice financial data modellers in the experimental design.

5 IMPLEMENTATION OF SEMANTIC MODELS AS A KNOWLEDGE REPOSITORY

5.1 Introduction

This chapter describes the implementation of an ontology and concept models as a knowledge repository for financial regulatory reports. Sections 5.2 provide an overview of the modelling and tooling choices made at the start of the implementation. The implementation approach described in Sections 5.3 to 5.5 emerged though an interactive process of identifying knowledge sources, knowledge acquisition and model implementation summarized in Figure 5.1.

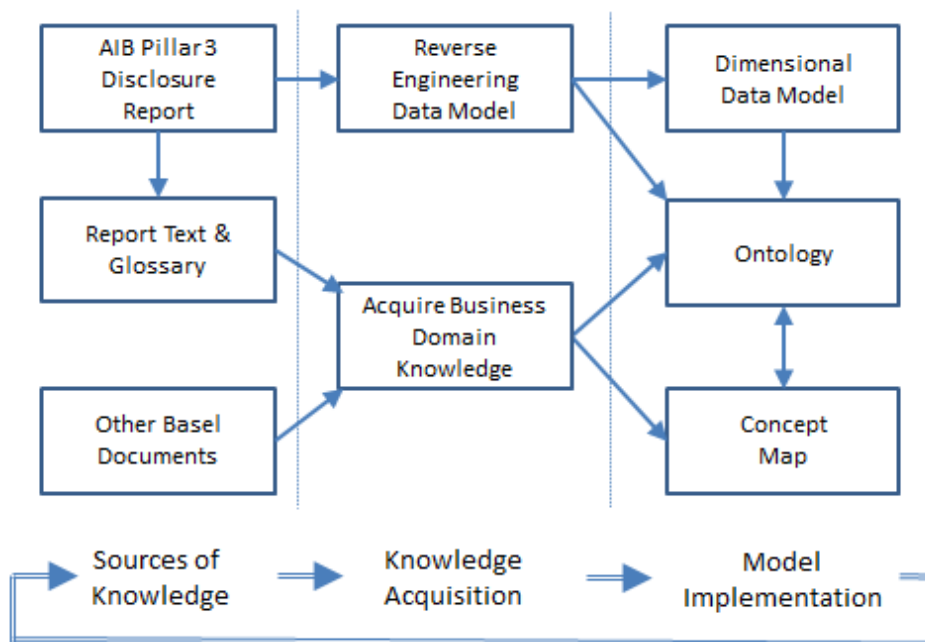


Figure 5.1 : Iterative Process of Model Implementation

The concept maps and ontology required a large amount of knowledge elicitation which resulted in a knowledge acquisition bottleneck. The implementation highlighted the importance of knowledge acquisition in development of semantic models and reinforces the similarity between semantic modelling and Nonaka's (2000) SECI process described in Chapter 3. An evaluation is presented in Section 5.6 that compares this implementation with the literature review and interviews in Chapter 2, 3 and 4. A short video that provides a narrated overview of the implemented concept map and ontology is provided in as part of the supporting material

5.2 Model and Tool Selection

Background research identified existing financial ontologies such as FIBO that could be have been used as a starting point for the experimental artefacts or that algorithmic approaches could be used to generate an ontology from a data model. It was decided to create the models from scratch using the knowledge modelling approach suggested by Milton (2007) as the creation of the models would provide an opportunity of observing the implementation challenges identified by the SMEs in Chapter 4. The approach is also an application of Smith's (2003) Oncologist's Credo in that the semantic modeller should gain an understanding of what they are modelling.

The evaluation of the types of semantic models used in the financial industry in Section 4.7 did suggest that a structured language such as SBVR could be used in this research instead of concept maps. It was decided to continue with concept maps as their visual structure and focus on end-user readability provided a greater contrast with the formal and technical structures of an ontology.

A short evaluation of the modelling tools identified in the background research was undertaken. The primary selection criteria were for tooling that would be easy for participants use during the experimental sessions. This supported the project objective to evaluate the usability of the semantic models and their implementation, and not for participants having to spend the time learning unfamiliar tooling. The secondary consideration was for a combination of tools that could be technical implemented and integrated in the timescale of the project.

IHMC Cmap Tools was selected as the tool for concept map development as it is recommended in the literature, was found to be straightforward to use and it allows for simple technical integration though the use of hyperlinks. The ontology tool Protégé was examined and found to have provide comprehensive modelling and integration functionality, but this tool was not familiar to experiment participants. The participants were familiar with IBM Information Governance Catalog (IGC) which provides a basic ontology support and has metadata integration functionality. IBM Information Data Architect (IDA) was chosen as the data modelling tool as it is both familiar to the participants and allows for integration between data models and ontologies in IGC.

5.3 Sample Reports and Data Model

The 2013 Basel Pillar 3 Disclosure from Allied Irish Bank (AIB) were selected as the sample business reports for the experiment because they related to financial industry regulatory reporting standards and include a number of tabular reports that provide the experiment with a good variety of business topics for knowledge modelling. This Basel reporting format is followed by other Irish bank such as Bank of Ireland and by a large number of regulated international banks. The Disclose report is prepared by the bank annually and contains quantitative information on capital requirements and risk management that the bank is obliged to provide to the Central Bank of Ireland. The AIB 2013 report contains a total of twenty eight individual tables or reports. The report also includes qualitative information in the form of commentaries and a description of risk management methods used by the bank. A copy of the AIB 2013 report is included as part of the supporting material.

Five of the twenty eight reports found in the document were selected for use in the experimentation and these are listed in Table 5.1. The five reports provided the scope of the knowledge for the semantic models that were to be implemented. The selection of the reports was made to provide one high level report that contained quantitative information that is then broken down in the other reports. The reports selected were not implemented in a reporting tool for the experiment.

Table 5.1 : Sample Reports and corresponding Data Model Fact Entities

Report in AIB Pillar 3 Document	Data Model Fact Entity
Table 2: Capital adequacy information – component of capital base (page 11)	Capital Adequacy Information Monthly Fact
Table 3a: Group capital adequacy information (page 12)	Capital Adequacy Information Monthly Fact
Table 4: Total exposures (EAD) by exposure class and related capital requirements (page 16)	Credit Exposure Weekly Fact
Table 5: Industry distribution of credit exposures (EAD) – Standardised Approach (page 19)	Credit Exposure Weekly Fact
Table 6: Geographic Distribution of credit exposures (EAD) – Standardised approach (page 21)	Credit Exposure Weekly Fact

While the focus of the research questions is on the knowledge repository, a data model was required for in order to ask a number of data modelling questions in the experiment described in Chapter 6. A dimensional logical data model was reverse engineered from the five selected reports. The data model follow the dimensional modelling approach recommended by Kimball and Ross (2011) with the exception of an example of snow flaking introduced for an experimental question on the topic of data model correctness. The entity relation diagram of the data model is presented in Appendix D.

5.4 Concept Map Implementation

The concept maps were developed by first making a list of the concepts used in the five reports and then embarking on an iterative knowledge elicitation of the relevant business domain knowledge. The result of the knowledge acquisition was the development of three concepts maps on the topics of *Basel II Capital Adequacy Framework*, *Credit Risk* and *Capital Requirements*. The concept maps were developed using the approaches recommended by Novak and Cañas (2006) and Milton (2007). The *Basel II Capital Adequacy Framework* concept map shown in Figure 5.2 was created as the entry point for the experiment and to provide a business context for the detailed concepts represented in the *Credit Risk* and *Capital Requirements* concept maps. All three concepts maps are presented in Appendix E and the source files are provided in the supporting material.

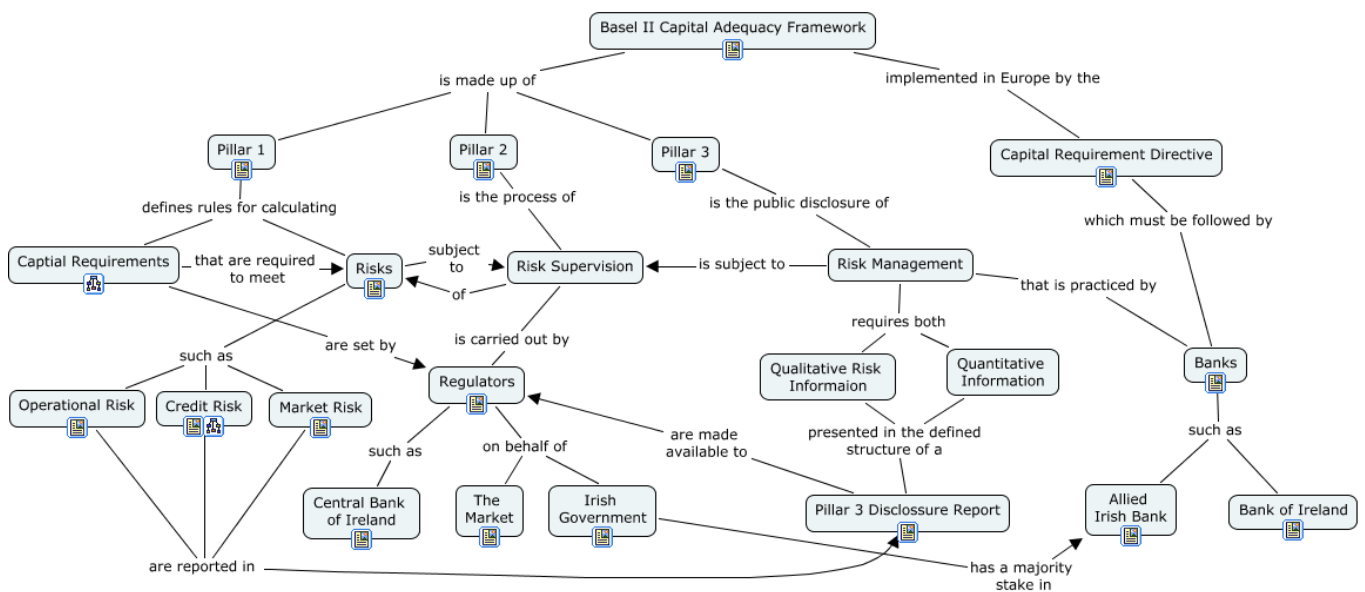


Figure 5.2 : Concept Map for Basel II Capital Adequacy Framework

The concept map implementation encountered some of the challenges of the knowledge acquisition bottle neck described in Section 3.3. The initial approach was to start to the elicitation plan by asking subject matter experts to provide a verbal overview of Basel regulatory reporting. However these experts were not available to participate in the project at the time of the implementation and so the knowledge elicitation had to rely on documentation. A variety of documentation sources were used including the Basel II: International Convergence of Capital Measurement and Capital Standard documentation (Basel Committee on Banking Supervision, 2006), commentary in the AIB Pillar 3 Document (AIB, 2013) and a guide on the calculation of capital ratios (Reserve Bank of New Zealand, 2007). This led to the narrow bandwidth challenge of having to read a large volume of dense technical financial texts while not having the general financial and regulatory knowledge to prioritise the acquisition or identify the core concepts. At the same time a significant latency of acquisition challenge emerged as the understanding of each financial concept invariably led to the need to acquire an understanding of many more related concepts.

The bottleneck was escaped by adopting a clear scope for the concept map that would provide a finishing line for the implementation. The scope of the concept maps was limited to the high level concepts for the selected reports and associated commentary from the AIB Pillar 3 document. It was decided that detailed calculation definition or complete classification hierarchies would be include in the ontology and would not be included in the concept maps. At this point the concept map and ontology started to be developed in parallel. The adoption of the scope concentrated the implementation on ensuring that combined semantic models contained the tacit knowledge required to read the five reports..

The final stage of the concept map implementation involved testing the knowledge with draft versions of the questions to be used in the experiment. This testing identified that the concept maps did not provide sufficient context to an individual who was unfamiliar the financial industry - one of the risks identified in Section 4.7.2. The high level *Basel II Capital Adequacy Framework* concept map was refined include the intended audience of the reports, examples of Irish Banks and a clearer overview of the Basel Framework. The two detailed concept maps were simplified so that they presented the main concepts required to understand the reports. A number of detailed concepts and cross-links were removed as they were cluttering concept maps and made

it difficult to read. Hyperlinks between the concepts map were added as an implementation refinement to make it easier to navigate from a concept to the related entry in the ontology. A HTML version of the concept maps were generated for the experiment as this allowed the semantic models to be available to the participants in a web browser rather than the unfamiliar IHMC Cmap Tools. The HTML for the concept maps is provided in the supporting material.

5.5 Ontology Implementation

The development of the ontology in IGC was undertaken in parallel with the development of concept maps. The fundamental unit of an ontology in IGC is the *term* which is analogous to a concept or object. An example term entry from the ontology is shown in Figure 5.3. The list of terms included in the ontology is provided in the Appendix F and as part of the supporting material.

Term Details

Edit Add to Collection... Delete

Tier 1 Capital

A measure of a bank's financial strength defined by the Basel Accord.

Long Description

Tier one capital is capital which is permanently and freely available to absorb losses without the bank being obliged to cease trading. An example of tier one capital is the ordinary share capital of the bank. Tier one capital is important because it safeguards both the survival of the bank and the stability of the financial system.

It captures core tier 1 capital plus other tier 1 securities in issue, but is subject to deductions relating to the excess of expected loss on the IRBA portfolios over the IFRS provision on the IRBA portfolios, securitisation positions and material holdings in financial companies. [Show Less](#)

Parent Category [Capital Adequacy](#)

Labels (1) [calculation](#)

Status Standard

▼ General Information

Calculation

Tier 1 Capital = (Paid Up Share Capital + Disclosed Reserves) - (Regulatory Tier 1 Adjustments + Supervisory Tier 1 Adjustments)

The following adjustment are applied by AIB

Regulatory Adjustments

- Defined Pension Scheme Provisions
- Intangible Asset such as goodwill

Supervisory Adjustments

- Financial Investments that are unconsolidated
- Securitisation

[Show Less](#)

Is Modifier False

Type Primary

▼ Associated Terms (5)

Is a Type Of (1) [Regulatory Capital](#)

[Expand hierarchy](#)

Has Types (2) [Disclosed Reserves](#)

[Paid-up Share Capital and Common Stock](#)

Synonyms (1) [Core Tier 1 Capital](#) / [Capital Adequacy](#)

Related Terms (1) [Core Tier 1 Ratio](#) / [Capital Adequacy](#)

Figure 5.3 : Ontology Term for Tier 1 Capital

A list of relevant business terms and definitions were extracted from the AIB Pillar 3 document and loaded into IGC to create a basic unstructured glossary. This glossary was extended to include a term for each of the data fields in the five reports and for each of the attributes in the data model. The basic glossary was refined into ontology through a combination of applying structure and improving text in the descriptions. The term names and descriptions were reviewed to ensure that they did not conflict with the content of the concept maps. This resulted in the refinement of some ontology terms names to ensure that each term has a unique identifiable name. The initial descriptions extracted from AIB Pillar 3 glossary were simplified and augmented with explanatory text from body of the document, BCBS Basel II documentation and equivalent glossary from the Bank of Ireland Pillar 3 Disclosure document². Care was taken to ensure that the short description were easy to understand because some of the explanatory text was detailed and assumed banking knowledge. Common abbreviations for terms were added the abbreviation property supplied by the IGC tool.

An initial structure was given to the glossary of terms by grouping them into two high level categories of *Credit Risk* and *Capital Adequacy*. This provided an alignment with the concepts in the detailed Credit Risk and Capital Requirements concept maps. Structure was added between the terms by using a combination of the basic ontology relationships of *Is a Type Of*, *Has Types*, *Synonyms* and *Related Terms* available in the IGC tool. Examples of these relationship types are show at the bottom of Figure 5.3 in the section labelled Associated Terms.

The *Is a Type Of* relationship was used to indicate that that one term is child, sub-type or component of another term. This relationship is the inverse of the *Has Types* relationship. For example, *Tier 1 Capital* is a component of *Regulatory Capital*, and in turn has the subcomponents of *Disclosed Reserves* and *Paid-up Share Capital*. The parallel development of the ontology and concept maps resulted in an iterative validation of consistency of the hierarchical structures used in both models. This is illustrated in Figure 5.4, where the term type hierarchy below the ontology term *Regulatory Capital* is show to be equivalent to the conceptual hierarchy below the

² <https://www.bankofireland.com/fs/doc/publications/investor-relations/boi-pillar-iii-disclosures-2013-11-15-28-03-14.pdf>

Capital concept in the concept map. This use of the *Is a Type Of* relationship for a wide range of ‘is part of’ relationships is perhaps overloading the intended use of the tools functionality, but the IGC tool does not support the definition of custom relationship types available other tools such as Protégé.

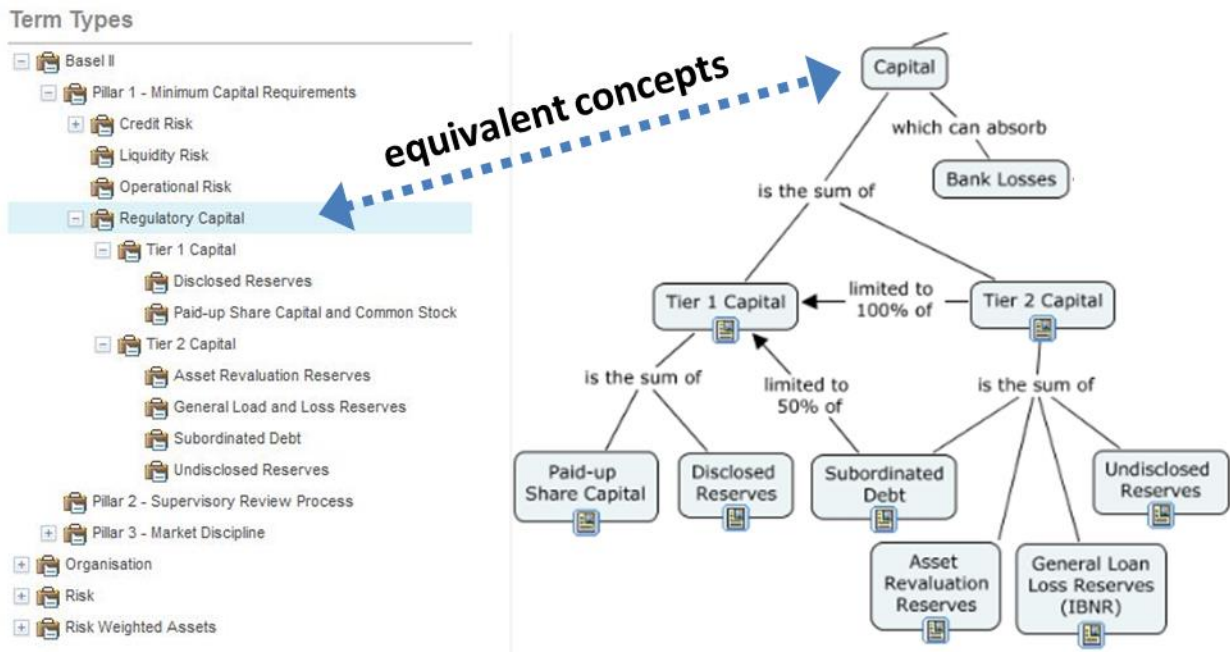


Figure 5.4 : Equivalency of Hierarchies in Ontology and Concept Map

The *Synonym* relationship was used to identify where two or more terms referred to the same business concept or report element. For example, because the terms *Tier 1 Capital* and *Core Tier 1 Capital* were used interchangeably in the sample reports and Basel documentation, a synonym relationship was created. To avoid confusion, one of the synonymous terms was used as the main term for this concept, in this example *Tier 1 Capital* was chosen as the main term.

The *Related Term* relationship was used as a generic method of identifying that one term was either referred to in the description or otherwise conceptually related to another term. The knowledge elicitation identified a requirement to represent the calculations in the ontology. Calculation descriptions were added by creating a custom property to specify the pseudo code for the calculations, and example of which is shown in Figure 5.3. Other ontology terms that are involved in the calculated were linked to the terms using the *Related Term* relationship. A further custom property was added to hold the risk weighting for standardised approach classifications.

The ontology and tool was configured to refine the navigation of the content. The advanced search options in the tool were configured so that text based searches would include a combination of the term names, short description, abbreviations and long descriptions. For example, Figure 5.5 shows the results of searching for *Capital Ratio* results in all terms that include the words *Capital* or *Ratio* in their name or descriptions.

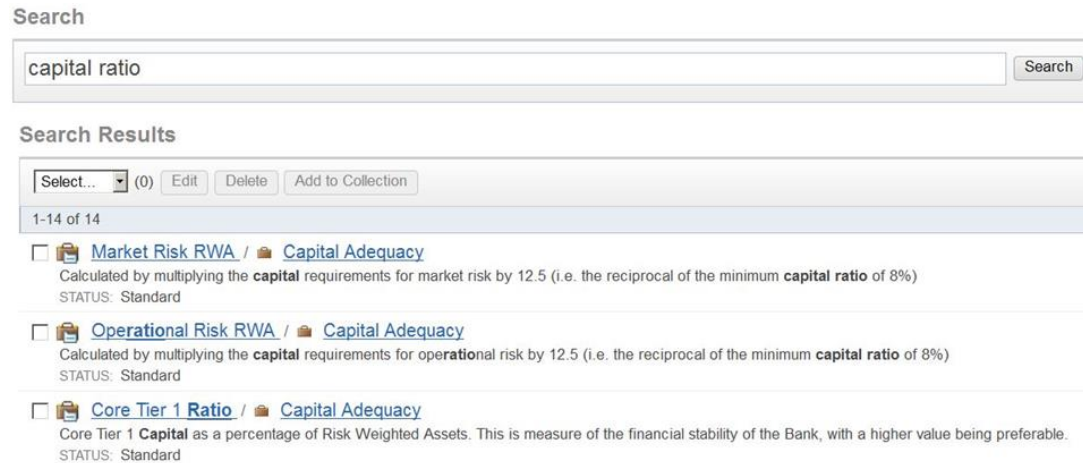


Figure 5.5 : Result of Searching Ontology for Capital Ratio

The final step of implementing the ontology was to create a relationship between terms and relevant entities or attributes in the data model. Making an explicit link from the data model to ontology was of interest as it was expected that a data modeller would find this useful way of navigating between the semantic models. The metadata of the dimensional logical data model was imported into the IGC tool which allowed for the linking of ontology terms to attributes using the *Associated Asset* relationship type. For most ontology terms there was an easily identifiable entities or attribute, for example the term *Paid-up Share Capital and Common Stock* was assigned to the attribute *Share Capital* in the *Capital Adequacy Information Monthly Fact* entity.

A different approach was used when mapping the classification values used in the report *Table 4: Total exposure (EAD) by exposure class*. This report breaks down the measure of exposure of default by the different classification used in the *Standardised and Foundation* approaches to credit risk. Each of the classification value used in the report has been represented as a term in the ontology and related to the credit risk approach using the *is a type of* relationship, for example Figure 5.6 shows an extract of the term type hierarchy that includes all the *Standardised Approach Classification Values*. Each of these classification values would be implemented as instance of the dimensional entity *Regulator Exposure Class* rather than an attribute in its own right

Figure 5.6 shows that the term *Real Estate Collateral* is associated with *Exposure Class Name* attribute in the *Regulator Exposure Class* entity.



Figure 5.6 : Example of Relating an Ontology Term to Data Model Attribute

5.6 Evaluation of Implementation

This evaluation discusses implementation and knowledge modelling described in this chapter. It contrasts the implementation experience of this research with similar findings from the literature review and the subject matter expert interview in Chapter 4. The following sub-sections address research questions 3 and 4 that were introduced in Section 1.5.

5.6.1 Implementing Semantic Models as a Knowledge Repository

The implementation of the semantic models for this research required significantly more knowledge elicitation than was originally planned for. This was a good example of the growing complexity challenge identified by the SMEs in Section 4.5.4 that warned that a semantic model quickly grows very quickly as the modeller seeks to represent how information elements are interpreted in many different ways. While the choice of the five sample reports did limit the scope of the concepts and terms in the models, each one of the elements had to be understood before it could be correctly

categorised in the structures of the ontology and the concepts maps. For example, it was not enough to simply identify that ‘Regulatory Retail’ was a type of ‘Standardised Exposure Classification’ – the term has to be understood to evaluate if it was a synonym of another term, how it was mapped to a data model elements etc. This complexity challenge could have been better addressed through a collaborative model development involving both semantic technical experts and business subject matter experts as suggested by Abi-Lahoud *et al.* (2014).

The implementation required an interactive development of the concept maps, ontology and drafts of the experimental questions. The availability of definitions in the ontology assisted in the correct understanding of concepts in the concept map. Adding a concept to either models typically triggered additions in the other model. Care was required to ensure that while there was an overlap in concepts, there was not an overlap in intended use of the models. The concept map was refined so that it provided a high-level overview with conceptual cross-links, while the ontology was used to focus on the detailed descriptions, classification and mappings to the data model. Sample questions were used to test the models to ensure that that it was possible to answer them using the content in the model. This identified that an ontology that simply describes complex information structure without providing good quality description is not useful as a knowledge repository. Concepts description need to include explanatory content and the semantic modeller cannot rely on the user to infer all meaning from the structure of the ontology.

The tooling choices described in Section 5.2 helped mitigate the implementation challenges relating to tooling and model standards discussed Section 4.5.4. However, there were still choices to made in the terms of the appropriate representation of the calculations, relationships and synonyms even with the selected ontology tool. It was found that the IGC is quite limited in the types of relationships that are allowable between ontology objects when compared to tools such as Protégé. The choice to integrate the concept map and ontology using hyper-links was initially intended as a validation step to ensure that each concept has a ontological reference where possible. However it was quickly identified as being very useful functionality when the implementation was tested with sample tasks. The possibility of importing the concept map into the IGC metadata was explored but it was not practical in the time available.

Research Question 3: What are the considerations when implementing semantic models as a knowledge repository?

Evaluation: The implementation of semantic models based repository requires a combination of the technical and business domain expertise. The use of more than one semantic model type encourages the adoption of an iterative implementation approach that emphasises cross-model validation of the knowledge content. While the choice of modelling tool can restrict the manner in which knowledge is represented care must be taken that the focus of the implementation is the provision useful knowledge for the end-user and not on unnecessarily complex modelling.

5.6.2 Comparing Semantic and Knowledge Modelling

The implementation of concept maps and ontologies described in this chapter allows for comparison of semantic modelling and knowledge modelling. While Chapter 2 discussed some of the differences between semantic and knowledge model types, the practical experienced of implementing an integrated set of models identified a lot of similarity in the modelling process. The requirement for both the semantic and knowledge modeller to have a good understanding of the knowledge they are modelling was discussed in Section 5.6.1.

Ontologies such as FIBO are intended to be both machine readable and human readable, but the requirement for human users to have the skills to consume content in technical format such as OWL or UML is a barrier to human understanding. In contrast knowledge models such as concepts maps are created for primarily for human consumption. While there is a basic format and guidelines for the creation of concepts maps, the knowledge modeller has significant freedom how they use the modelling techniques when compared to modelling within a formal ontology. This difference can be described as one of communication versus specification. Each concept map is an individually crafted model with the purpose of communicating knowledge on a specific topic to other humans, and the modeller is in control of the format and the manner in which it will be displayed to users. Ontologies are a formal specification of the allowed relationships between objects and their properties and are intended to be expressed in a limited of syntax for consumption by both human and machine.

Using Moody and Shanks (2003) model quality factors; knowledge modelling emphasises the factors of understandability and *simplicity* of the model, while semantic modelling emphasises *correctness* and *implementability*. Examples of these factors when creating this ontology included ensuring the completeness of classification values, a consistent approach when mapping to data model attributes and correct application of the relationship types provided by the IGC tool. In contrast the concepts maps were simplified to make them easier to read and the predicates were tested to ensure they formed understandable sentences. It should be noted that some of the simplicity of concept maps was achieved because it was possible link the concept map to ontology terms that contained detailed description of calculation examples.

There is an expectation in the financial industry that a semantic model must be capable of helping define the business rules logic for applications; providing the bridge for business and IT communication; and form the basis of information governance. The modelling experience described in this chapter shows that it is difficult to use a single model format to represent the specialist business knowledge and information structures that could be required for all these uses. The facility to link the ontology and concept models allows the modeller to take advantage of the strengths of the different modelling tools by having multiple representations of the same concept. Figure 5.7 illustrates that the abstract knowledge in the repository can viewed as have been implemented in a number of humans and machine readable formats. This figure is similar to Figure 2.9 which illustrate how multiple document formats can be used to express a single ontology represented in OWL

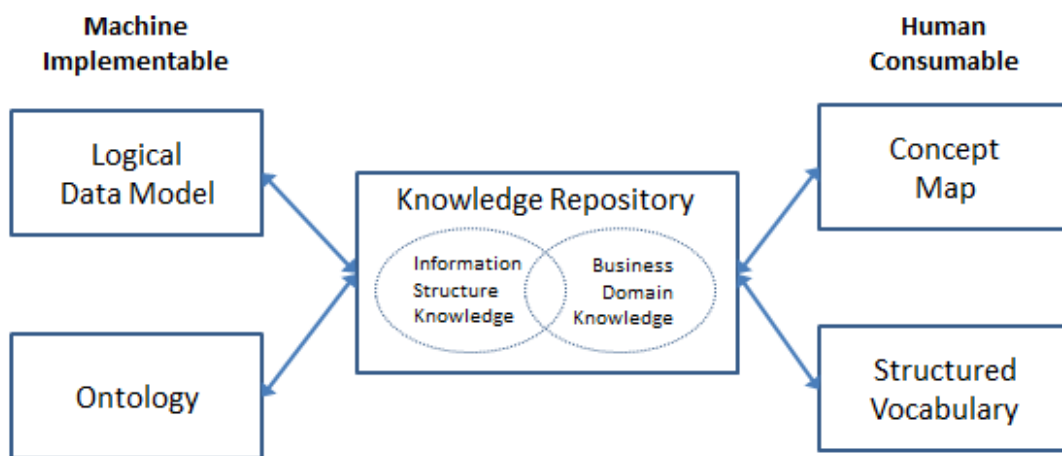


Figure 5.7 : Human Consumable and Machine Implementable Knowledge

Research Question 4: What is the difference between semantic modelling and knowledge modelling?

Evaluation: Semantic modelling utilise formats that emphasise model correctness and machine implementation, while the formats used by knowledge modelling emphasise human communication and learning. Both types of modelling require modellers who have an understanding of the knowledge that they are representing. There is trend in the financial industry for semantic models that are both formally correct and understandable by both business users and technical modelers. This suggests that a combination of semantic and knowledge modelling techniques should be applied in the implementation of knowledge repositories.

5.7 Conclusions

This chapter described the implementation of the concepts maps and ontology were implemented as a knowledge repository to support the use of five sample Basel regulatory reports. The approach outlined in Section 5.1 was followed successfully during the implementation but there was a need to carefully maintain the scope of models so that the timelines of the project were met. Even with a controlled scope, the creation of the ontology and the concepts maps described in Sections 5.4 and 5.5 required a significant amount of effort to both implement the technical artefacts to the acquire the financial knowledge represented in the models.

The evaluation of the implementation concluded that the use of a combination of both ontology and concept maps supported the acquisition of knowledge by the modeller and encouraged the validation of the knowledge in the models. The implementation used modelling tools that are representative of those used in the financial industry and education. While the choice of modelling tool did limit the ways in which the knowledge could be represented, they were sufficient to explore the semantic and knowledge modelling techniques that can applied to a knowledge repository. Chapter 5 presents the design and results of the experimental use of this repository by data modellers who were tasked with answering question about the sample regulatory reports.

6 EXPERIMENTAL USE OF SEMANTIC MODEL BASED KNOWLEDGE REPOSITORY

6.1 Introduction

This chapter describes the design and result of an experiment where data modellers were asked to complete a series of questions that required them to use knowledge represented in the ontology and concept models described in Chapter 5. Sections 6.2 and 6.3 describe the experimental approach and the design of the questions chosen to support the evaluation of the research goals. Section 6.4 describes the environment where the experimental sessions were carried out and gives a short profile of the five data modellers who participated in the study. Section 6.5 and 6.6 summaries the quantitative results and the participant's qualitative feedback on the semantic model content and implementation is provided. An evaluation and discussion of the experimental results is provided in Section 6.7.

6.2 Experimental Approach

The objective of the experiment was to evaluate the ontologies and concept maps whose implementation is described in Chapter 5. The interviews in Chapter 4 identified that individuals frequently preform the role of both business analyst and data modeller at the same time, inhabiting the middle ground between business users and IT systems. The experimental approach taken was to measure how useful the semantic models were for answering pairs of questions that span the business and IT gap. In the experiment the data modeller has the choice to use either or both the concepts maps and ontology to answer the questions. The design of the questions required the data modeller engage with the knowledge represented in the repository. The knowledge needed to answer the questions was represented in both models to a support a comparison of the usefulness of models types when answering the questions. The approach is illustrated Figure 6.1 which shows examples information structure and business domain questions relating to Tier 1 Capital Ratio.

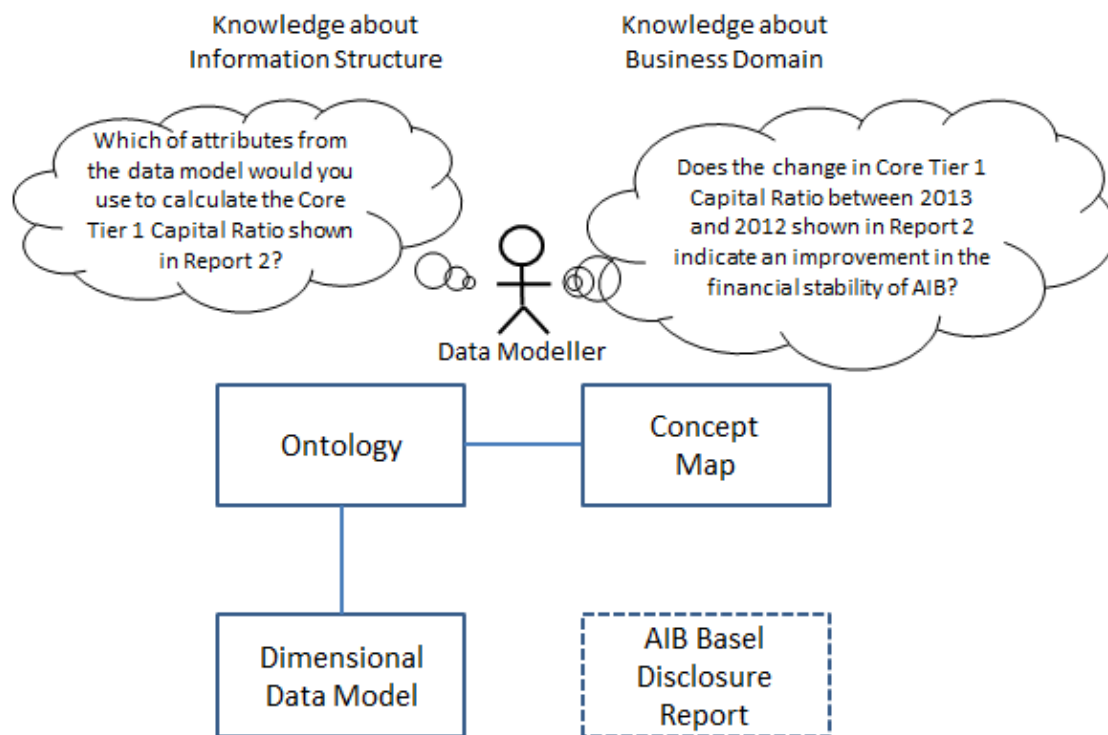


Figure 6.1 : Example of Experimental Questions

The experiment was designed to allow for comparisons with the three studies in the literature that evaluated the usefulness of semantic models to data modellers when completing modelling tasks. Osman's *et al.* (2011) study of the usefulness of ontologies asked novice data modellers to rate their modeller's perception of the usefulness of the ontology when completing a modelling. The modellers were also asked to describe any difficulties they faced completing the task or using ontology using open ended questions). Vieritz's *et al.* (2013) assessment of an online learning environment that contained a knowledge model asked both students and teachers to rate the knowledge management system both in terms of their perception of how effective it was for both learning goals and teaching goals. The Data Model Quality Factors described by Moody and Shanks (2003) provide a framework for asking the data modeller to evaluate the both model implementation and usefulness through a combination of open questions and feedback on the practical use of the models.

The minimum experience for individuals selected to participate in the study was one year of applied data modelling experience in any industry or domain. Participants with more than 2 years of experience in financial industry would be considered expert financial data modellers. The experiment was designed to be completed by one data modeller at a time with the knowledge repository made available on a laptop. Each

experimental session was planned to take approximately one hour broken down into three stages;

1. Introduction to the study and short tooling tutorial (10 minutes)
2. Completion of the ten tasks in the questionnaire (30-40 minutes)
3. Feedback questions and discussion (10 minutes)

The short introductory demonstration was required to ensure that the participants would be familiar with the tool functionality and would be able to navigate and search both the ontology and the concept map. The questions were representative of the tasks typically encountered by data modellers while also having an expected completion time of approximately three to four minutes. The short completion time was used to give participants a realistic chance being able to attempt all ten questions.

6.3 Design of Questions and Metrics

6.3.1 Questions to Require Business or Information Knowledge

The experiment questions required either business or information knowledge to answer. A total of ten questions were selected with each type of knowledge having five associated questions. A question that related to the understanding of a banking concept or the interpretation information in a report was classified as requiring business domain knowledge. A question that related to the data model, classification hierarchies or performing calculation was classified as requiring information structure knowledge. Examples of both types of the questions are provided in Table 6.1.

Table 6.1 : Comparing Business and Information Questions

Business Domain Questions	Information Structure Questions
<p>Understanding a business concept</p> <p>Example Question: What are Risk Weighted Assets (RWA)?</p>	<p>Understanding the classification and calculation of information</p> <p>Example Question: If a €1,000 loan is past due payment by more than 90 days, what would its Risk Weighted Asset value be using the Standardised Approach?</p>
<p>Interpreting information in a business report</p> <p>Example Question: Does the change in the bank's Core Tier 1 Capital Ratio between 2013 and 2012 indicate an improvement in the financial stability of AIB?</p>	<p>Identifying which attributes in a data model support a business report.</p> <p>Example Question: Which of attributes from the data model fact entities would you use to calculate the Core Tier 1 Capital Ratio shown in Table 2?</p>

The level of difficulty of the questions was designed so that participants would be challenged when completing the tasks but would still be at a level where all participants should be able to attempt the majority of the tasks. The questions were tested to validate their classification and they could be answered with the content in the model. This testing identified a number of gaps in the model content and so became part of the iterative process of semantic model development described in Chapter 5 . The ten experiment questions are listed in Table 6.2 and the full experiment questionnaire is provided in the supporting material.

Table 6.2 : Experiment Questions

#	Experiment Questions	Knowledge Type	Difficulty
1	Which of the following is the primary audience of the Pillar 3 Disclosure Reports published by AIB?	Business	Easy
2	What is the target Core Tier 1 Capital Ratio set by the Central Bank of Ireland for AIB?	Information	Easy
3	What are the different types of risk that must be managed by AIB?	Business	Easy
4	What are Risk Weighted Assets (RWA)?	Business	Medium
5	In relation to Table 2, does the change in the bank's Core Tier 1 Capital Ratio between 2013 and 2012 indicate an improvement in the financial stability of AIB?	Business	Medium
6	What types of Risk Weighted Assets are used in the calculation of Capital Ratios?	Business	Easy
7	Which of attributes from the data model fact entities would you use to calculate the Core Tier 1 Capital Ratio shown in Table 2?	Information	Medium
8	If a €1,000 loan is past due payment by more than 90 days, what would its Risk Weighted Asset value be using the Standardised Approach?	Information	Medium
9	In Table 4, would you include a mortgage to a retail customer in the calculation of total Retail Exposure in the Standardised Approach to Credit Risk?	Information	Hard
10	The information in Table 6 is supported by the entity Credit Exposure Weekly Fact. Is this entity modelled correctly to support the calculation of 'average exposures over period' if corporate customers can change country of operation in the middle of the year? If not, why not?	Information	Hard

6.3.2 Metrics of Model Usefulness

Metrics were used to measure and observe the overall and relative usefulness of the ontology and concept maps to the data modellers. The quality framework suggested by Moody and Shanks (2003) emphasises the use of a small number of quantitative metrics when attempting to measure model quality. On completion of each question the participant was asked to identify which of the models was most useful for completing the question and to rate how useful they found that models in answering the question. A five point rating was chosen to allow for comparison with the Osman’s *et al.* (2011) study. The three metrics chosen are listed in Table 6.3.

Table 6.3 : Metrics for Model Usefulness

Metric	Measure	Collection Method
Most Useful Model	Single choice of Concept Map, Ontology, Both and Neither	Participant was asked ‘Which semantic model was most useful for this question?’
Helpful Rating	Scale of 1 to 5, where 1 is not helpful and 5 is very helpful	Participant was asked ‘Please rate how helpful this model was in answering this question.’
Answer Correctness	Yes or No	Answer provided by the participant evaluated after the experiment session.

The selection of a simple measurement approach also had the benefit of being easy for the participant to understand and complete. An example of the questionnaire format used for all ten questions is illustrated in Figure 6.2.

Question 5.

In relation to Table 2, does the change in the bank’s Core Tier 1 Capital Ratio between 2013 and 2012 indicate an improvement in the financial stability of AIB?

5.1 Answer

- Yes
- No

5.2 Which semantic model was most useful for this question?

- Concept Map
- Ontology
- Both
- Neither

5.3 Please rate how helpful this model was in answering the question

- 1
 - 2
 - 3
 - 4
 - 5
- Not Helpful Very Helpful

Figure 6.2 : Example of Format Used in Experiment Questions

In addition to the three qualitative metrics on usefulness, a set of five questions were designed to elicit qualitative feedback from the participants once they had completed the tasks. The intention of these questions was to get the participant to reflect on the overall usefulness of the knowledge in the models and to share their thoughts on how the implementation could be improved. While the questions were included in the questionnaire it was intended that they would initiate a discussion with the participant rather than providing a simple yes or no answer. The questions were designed to allow for a comparative evaluation with the qualitative results reported by Osman *et al.* (2011), Vieritz *et al.* (2013) and Moody and Shanks (2003).

6.4 Overview of Experimental Sessions and Participants

The experiment was conducted with five data modellers who are members of a software organisation that is engaged in the development of data models and process models specific industries including banking, insurance and healthcare. The experiment was conducted in the participant's workplace with each session taking approximately one hour. Each participant was provided with a paper copy of the experiment questionnaire, the logical data model and the five sample reports from the AIB 2013 Disclosure document. The semantic models were made available on a laptop running a running the IGC and IDA tools in a virtual machine. The participants accessed the concept maps and the ontology using the Firefox web browser.

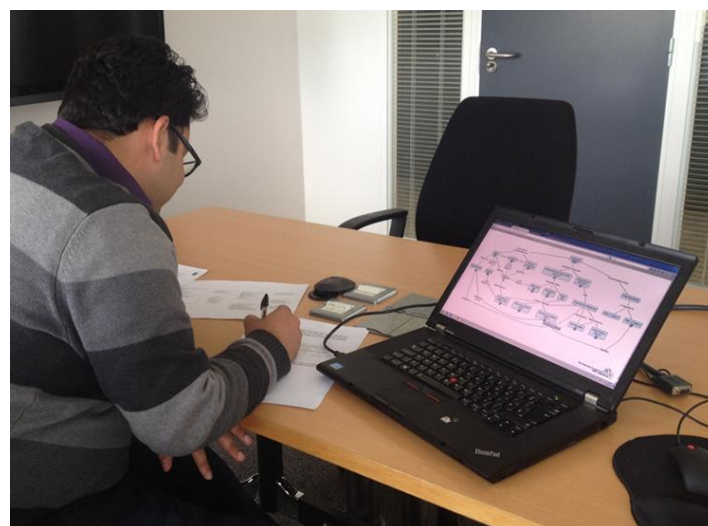


Figure 6.3 : Photo of Participant Performing the Experiment Tasks

Each of the participants was asked to describe their current role and to indicate the number of years of experience working with financial or banking data models. While all of the participants had at least two years of general data modelling experience, only two of the participants had the more than two years of experience in financial industry for them considered as expert financial data modellers for the purposes of this experiment.

Table 6.4 : Profile of Experiment Participants

Participant	Current Role	Experience with Financial Data Models
Expert 1	Data Modeller and Business Analyst	2 to 5 years
Expert 2	IT Architect	5+ years
Novice 1	Data Modeller	0 to 2 years
Novice 2	Data Modeller and Business Analyst	0 to 2 years
Novice 3	Data Modeller and Business Analyst	0 to 2 years

Each participant was given a verbal introduction to the experiment and was guided through a short tutorial that demonstrated how to open the concept maps and the ontology in the web browser. The script for this introduction is provided in the supporting materials. At this point the participants were encouraged to spend a few minutes exploring the model content before attempting the questions. It was emphasised that the study had more interested in obtaining their opinions on how useful the semantic models were in completing the tasks, rather than measuring if they gave right or wrong answer to the questions. The experimenter sat beside the participant throughout the session so that the use of the semantic models could be overserved and recorded. This also facilitated discussion and feedback on the use of the models throughout the session.

6.5 Results from Experiment Questions

Each of the five participants was asked to complete ten questions and to identify and rate the model that they found most useful for each question. Three of the five participants completed all ten questions, while two of the participants completed the first eight questions. The participants who did not complete all questions were not able to do so due to the time restrictions on the experiment. The answers given by the

participants were reviewed after the session and marked as either correct or incorrect. The full result data set from the experimental questions and the observation made by the experimenter are provided in the supporting material.

The statistical analysis of the results in the following sections is limited to the use descriptive statics including the mean, mode and frequency distributions. It is not appropriate make any statistical inferences from the results due to the small number of data points and the likely inter-dependencies between the variables measured.

6.5.1 Summary of Quantitative Metrics

The combined participants completed a total of 46 questions. The most frequent response of participant’s responses to ‘Which semantic model was most useful for this question?’ was the ontology. The distribution of the *most useful model* metric for all questions is illustrated in Figure 6.4, which shows that while the ontology most useful, it was also frequently found to be useful in combination with the concept map.

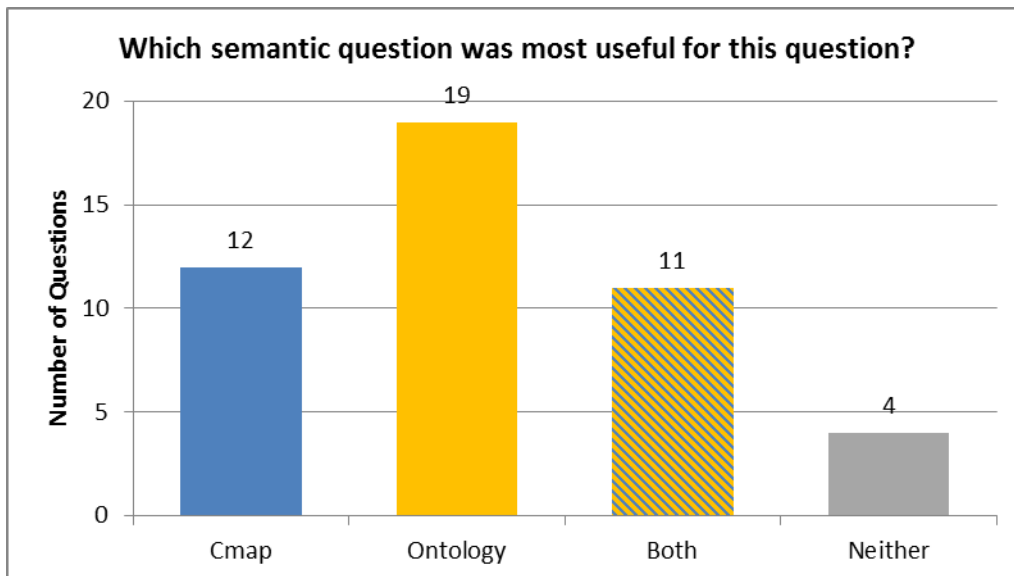


Figure 6.4 : Distribution of Most Useful Model for All Questions

The mean of participant’s responses to ‘Please rate how helpful this model was in answering the question’ across all questions was 3.8, which is between 3 (Neutral) and 4 (Helpful). The participants rated the semantic model as helpful or very helpful in 74% (34 of the 46) of attempted questions. The distribution of the helpful rating metric is illustrated in the Figure 6.5.

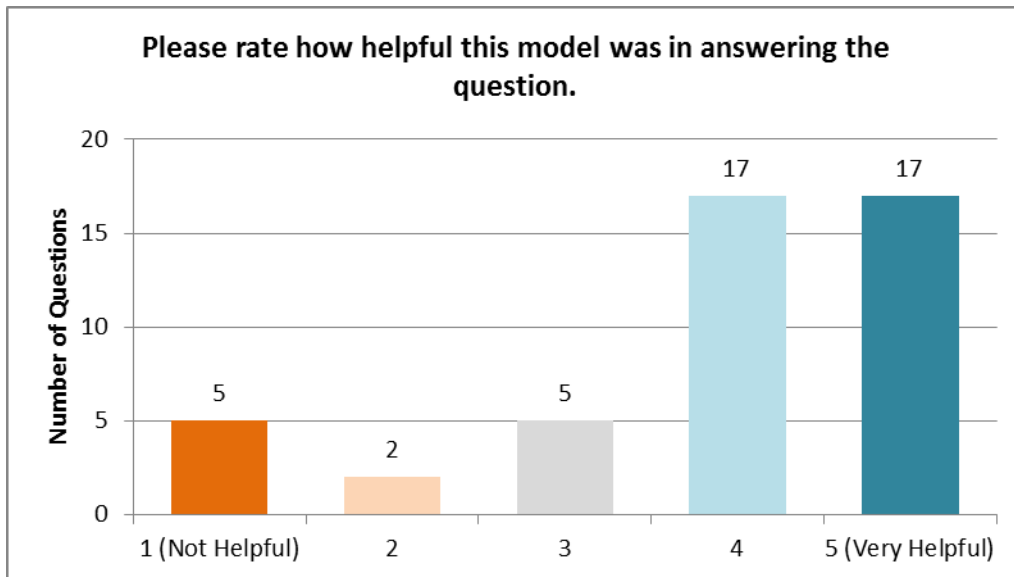


Figure 6.5 : Distribution of Helpful Rating for All Questions

A summary of the *correctness*, *most useful model* and *helpful rating* metrics for each of the ten experimental questions is provided in Table 6.5.

Table 6.5 : Summary of Experimental Question Metrics

Question	Question Attempts	% Correct	Most Useful Model ³	Helpful Rating ⁴
1	5	80%	Cmap	4.4
2	5	80%	Cmap	4.6
3	5	100%	Both	3.2
4	5	100%	Ontology	4.8
5	5	80%	Both	4.6
6	5	100%	Draw	4.4
7	5	20%	Both	2.2
8	5	100%	Ontology	4.2
9	3	0%	Ontology	4.0
10	3	67%	Neither	1.0
Total	46	76%	Ontology	3.8

³ Mode of participant responses to ‘Which semantic model was most useful for this question?’

⁴ Mean of participant responses to ‘Please rate how helpful this model was in answering the question’, where 1 is not helpful and 5 is very helpful.

6.5.2 Analysis of Answer Correctness Metric

The analysis of the non-attempts, correct and incorrect answers in Figure 6.6 shows a similar distribution between both expert and novice participants. This suggests that the novice's low level of banking regulatory domain knowledge was not a barrier to the effective use of the semantic models to complete the experimental questions.

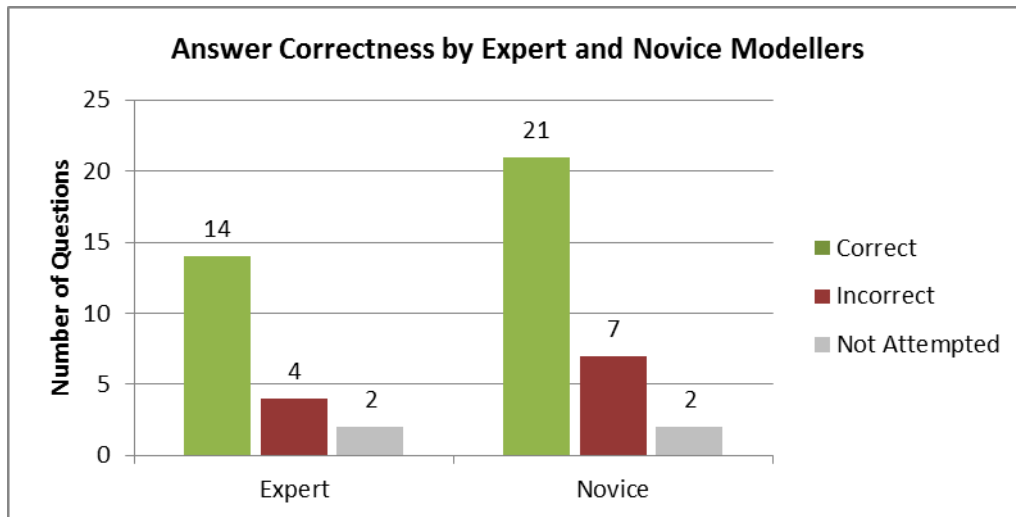


Figure 6.6 : Analysis of Question Correctness by Expert and Novices

The most frequent incorrectly answered were questions 7 and 9 which were answered incorrectly by more 50% of the participants who attempted the question. Both these questions required the participant to use Information Structure knowledge to complete the task.

Question 7 required the participant to identify the attributes in the data model required for a calculation of the Tier 1 Capital Ratio included in sample report Table 2. The one participant who correctly answered the question was an expert who used their existing knowledge to find the attributes that matched the report names. All other participants did identify some of the attributes by examining the calculation of Tier 1 Capital Ratio in the ontology but they all observed to be were unsure of the quality of their answer.

Question 9 was designed as hard question that that required a careful reading of both a classification definition and structure provided in the ontology. All Participants correctly identified the classification structure but then incorrectly identified the classification value. Participants were typically confident in their answer of this question and rated the ontology as being useful in completing the answer.

6.5.3 Analysis of Most Useful Model Metric

A distribution of *most useful model* metric categorised by questions that required business domain or information structure knowledge is illustrated in Figure 6.7. Both types and combinations of models were reported as equally useful when answering Business domain questions. The ontology was identified as being most useful for over 50% of the information structure questions.

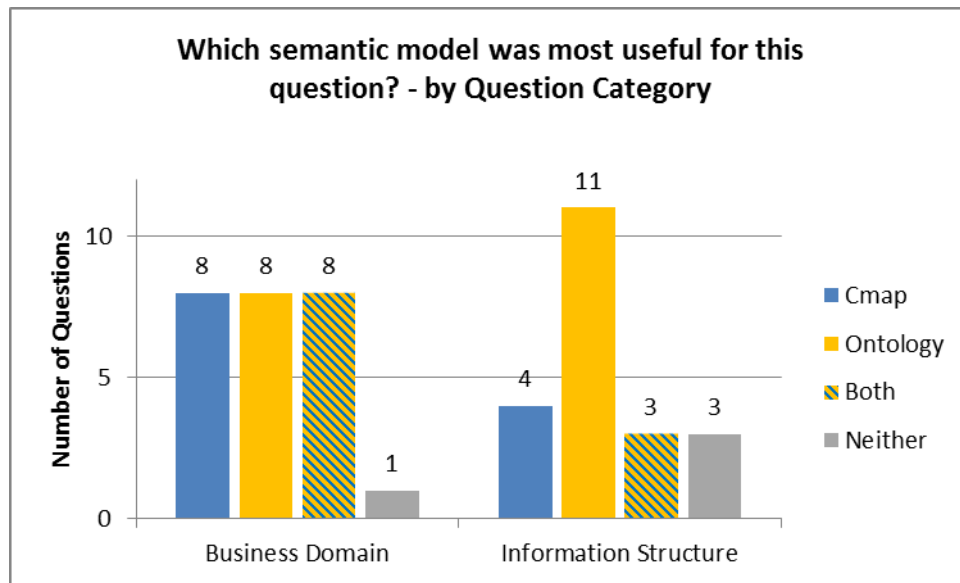


Figure 6.7 : Distribution of Model Usefulness by Question Category

Questions 8 and 9 were the two questions where the ontology was identified as the most useful by more than 80% of participants. The comparative usefulness of the ontology for Question 9 reflects the fact that the concept map did not contain all the required information to answer the question. Question 8 required the participant to identify a risk weighting for a loan and apply it to the calculation of risk weighted assets. The question was correctly answered by all participants and they were observed to all use a combination of ontology and concept maps when answering the question. However, the ontology contained a very clear example of the calculation and more than one participant remarked that the answer was simple once they identified the example in the ontology.

During the experiment sessions it was observed that some participants showed a preference for one model over the other. This observation is supported by the analysis of model usefulness illustrated in Figure 6.8 which shows a high the variation between participants.

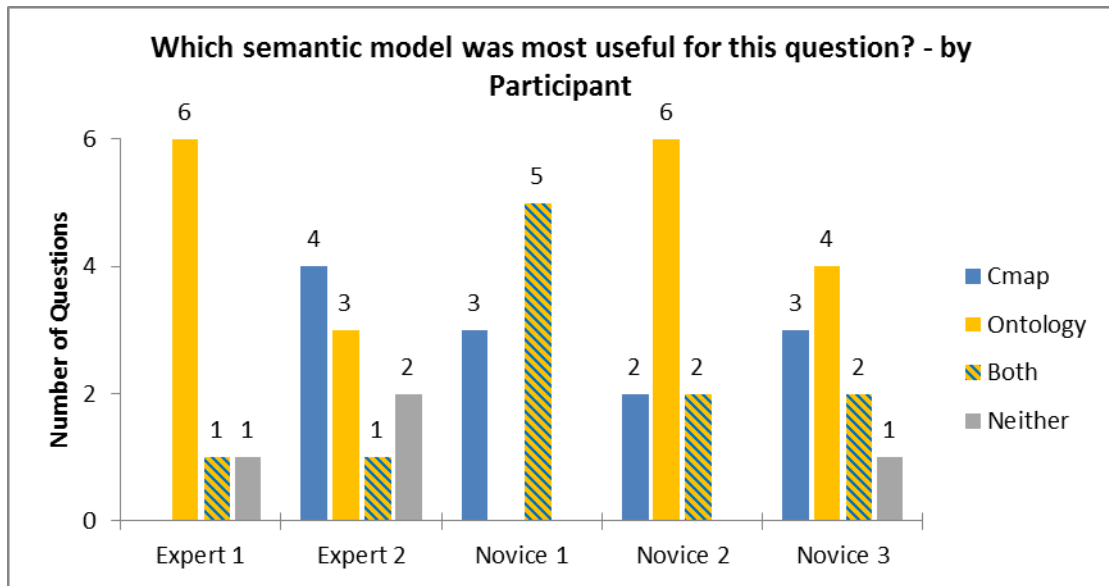


Figure 6.8 : Distribution of Model Usefulness by Participant

Expert 1 and Novice 2 were both observed to favour the ontology because they had a preference for navigating the content using the search functionality in the IGC tool. In contrast, Novice 1 studied the concept maps for approximately five minutes before starting the questions start and then went on to favour the concepts map alone or in combination with ontology as the most useful model.

A combined analysis of the *correctness* and *most useful model* metrics results in the distribution illustrated in Figure 6.9 This suggests the correctness of the participant’s answer is independent of the type of model they found useful for that question.

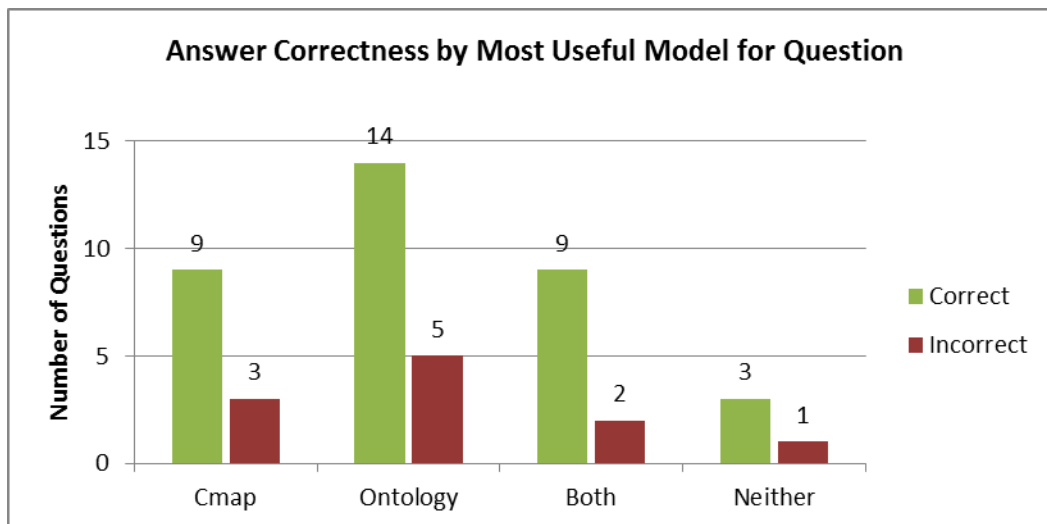


Figure 6.9 : Distribution of Answer Correctness by Most Useful Model

6.5.4 Analysis Helpful Rating Metric

The analysis suggests the models were more useful for business domain questions. The participants rated the sematic model as helpful or very helpful in 88% of questioned requiring business knowledge and in 62% of questions requiring information structure knowledge. The summary statistics of the *helpful rating* are provided in Table 6.6, and the distribution between question categories is illustrated in Figure 6.10.

Table 6.6 : Helpful Rating Summary Statistics

Question Category	Question Attempts	Mean Helpful Rating	St Dev Helpful Rating	% Helpful	% Not Helpful
Business Domain	25	4.3	0.9	88%	4%
Information Knowledge	21	3.3	1.5	57%	29%

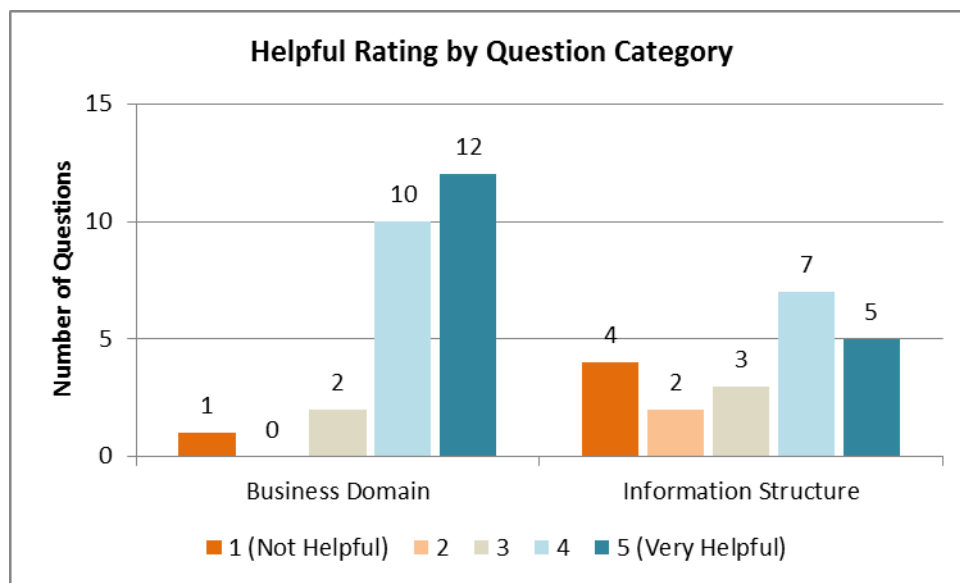


Figure 6.10 : Distribution of Helpful Rating by Question Category

The participants gave the models a helpful rating of 1 or 2 in only seven (15%) of the responses. The distribution of the helpful rating metric provided in Figure 6.11 show that four of these seven negative responses were responses were given where the participants found neither of the models useful.

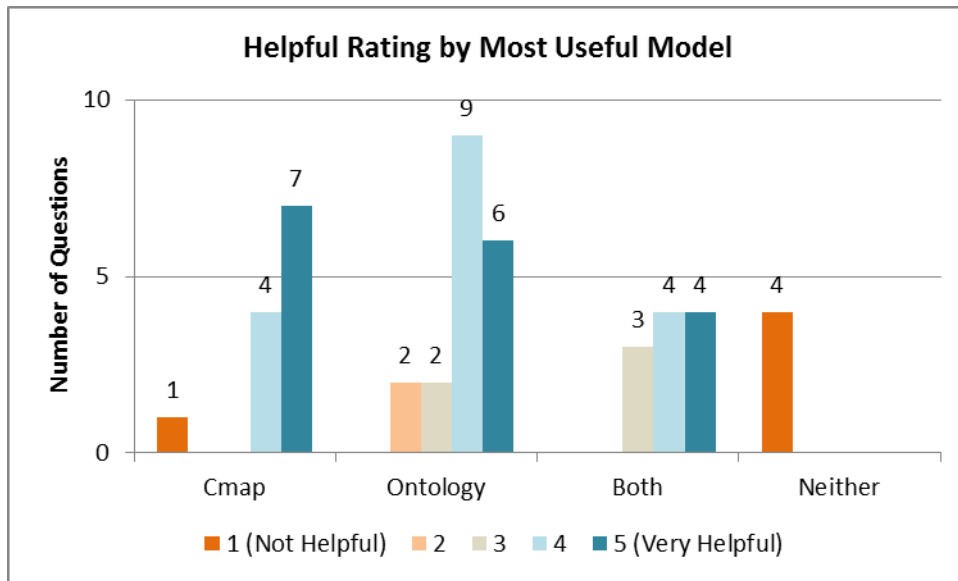


Figure 6.11 : Distribution of Helpful Rating by Most Useful Model

The distribution of the helpful rating metric provided in Figure 6.12 shows that participants frequently rated a model as helpful where they provided an incorrect answer, or vice versa.

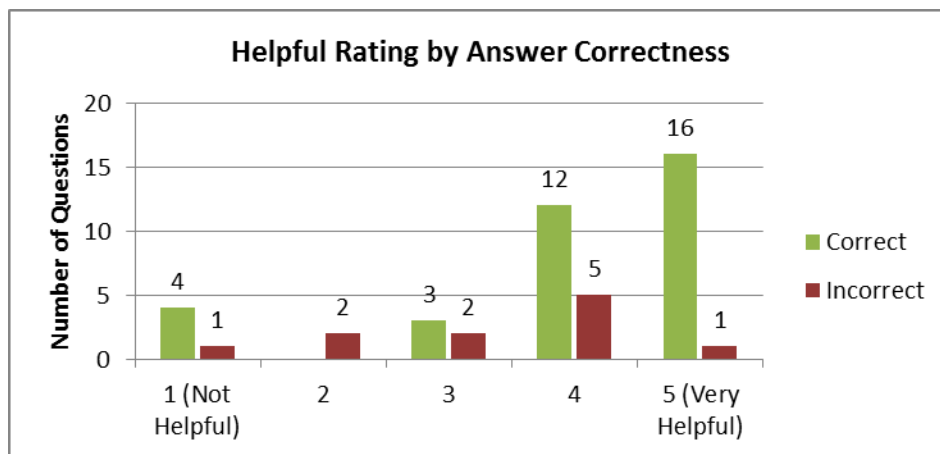


Figure 6.12 : Distribution of Answer Correctness by Helpful Rating

6.6 Results from Feedback Questions

After completing the tasks questions the participants were asked five feedback questions. The first two questions asked the participant if they would consider the ontology and concept maps to be a suitable tool for learning about financial regulations. The three remaining questions were used to elicit verbal feedback from the participants on ease of understanding the models, completeness of knowledge and model implementation.

6.6.1 Feedback Question 1 and Question 2

All participants indicated they would both use and recommend a knowledge repository comprising of an ontology and concept map.

#	Feedback Question	Yes	No
1	Would you use ontologies and C-Maps for learning about Basel Regulations if they were available and you had to perform financial data modelling	5	0
2	Would you advice novice financial data modellers to use this or a similar knowledge repository??	5	0

6.6.2 Feedback Question 3

Question: **Did you find the content in the models easy to understand?**

Participant	Comments
Expert 1	I prefer to use visual models but I did not have the time to read and digest the concept map. Would need an hour to digest the concept map content. Some of the questions were very specific, so the search functionality in the ontology was very appealing. Not sure if the questions were conducive for using the concept map.
Expert 2	Yes. The concept map is presented in a non-intimidating fashion. It took a while to get a handle on how I should use one model or another. I needed to get a better feel of what task is best suited to which type of model.
Novice 1	Yes. Concept map is good starting point for understanding the subject area.
Novice 2	Yes, but needed a combination of both models Concept map for structure and was easy to understand unless very specific questions. Ontology was good for descriptive questions and definitions
Novice 3	Yes, as a result of the C-Maps. The concepts jump out at you and are easy to navigate up and down. The link from concept map to Ontology is also important.

6.6.3 Feedback Question 4

Feedback Question: **Did the models provide you with the complete knowledge to answer the questions?**

Participant	Feedback
Expert 1	Think that the concept map is knowledge for business users. If I had more knowledge I may have used the concept map more
Expert 2	Almost. The combination of term-type hierarchy in the ontology for definitions and concept map to give context works well.
Novice 1	Yes. Detailed questions answered in the ontology and generic questions answered from the concept map.
Novice 2	Yes, 90%, all but question 10.
Novice 3	For most parts

6.6.4 Feedback Question 5

Feedback Question: **Did you think that any of the content was either incorrect for business or technical reasons?**

Participant	Feedback
Expert 1	Lack of experience with the tool was an issue for me. It was good that I would jump from the higher level concept maps to the ontology, but I would like to be able to go back the other way
Expert 2	Benefit of concept map is that it is informal, but some people prefer more structure. It was good that the concept map was organised in cells. The amount on each page seemed good.
Novice 1	The link to data model metadata was useful. More specific question needed the link from concept map to ontology. The concept map allows a mixture of relationship types. The ontology would be better if you were not constrained by the tool, and it allowed relationships with free description and direction.
Novice 2	Concept map could be colour coded or could use different shapes show instance of types. All the concepts being the same colour is boring, and the model does not look very useful until you start reading it. Included a legend in the C-Map to explain what the shapes, icons and arrows mean. Concept map to ontology links are very good.
Novice 3	Maybe some more example or definitions on the C-Map would have been good. Having a pop-short description when you hover over the concepts

6.7 Evaluation of Use of Knowledge Repository

This evaluation analyses the quantitative and qualitative results of the experiment presented in Section 6.5 and 6.6. The results are compared with similar results of Vieritz *et al.* (2013) and Osman *et al.* (2011). The following sub-sections addresses research questions 5, 6, 7 and 8 that were introduced in Section 1.5.

6.7.1 Did the semantic models provide useful knowledge to data modellers?

The approach taken to evaluate how useful the knowledge was to data modellers was to ask the participant to perform tasks that required the knowledge to interpret a financial regulatory report. The ‘usefulness’ of the knowledge was captured by both asking the participants to rate how helpful the model was and to measure how correctly they answered the questions. The participants answered 76% of the questions correctly which indicates that they were able to effectively apply the knowledge to the tasks provided. Interestingly the participants frequently rated a model as helpful or very helpful when they provided an incorrect answer,

The mean helpful rating across all the experiment questions was 3.8, which is between 3 (Neutral) and 4 (Helpful). This is comparable with a mean rating of 3.5 from Vieritz’s *et al.* (2013) study where students were asked to rate the usefulness of an e-learning environment that utilised a knowledge map on a similar five point scale. The models were rated as helpful or very helpful in 34 of the 46 (74%) of attempted questions. This contrasts with the results of Osman’s *et al.* (2011) study on use of an ontology in a learning environment where only 29% of novice data modellers to perceived the ontology as ‘useful’.

The qualitative evaluation used three of the factors from the data model quality framework described by Moody and Shanks (2003).

- *Understandability.* Four of the five participants found the content of the models easy to understand with the concept map being identified as aiding comprehension, but it was remarked that there was not sufficient time to digest the content during the session.
- *Completeness.* The participants commented that the models almost provided the complete knowledge required for the tasks but there were some small gaps.

- *Correctness.* While all five of the participants found the models correct e.g. they did not identify any incorrect knowledge represented in the model, it should be noted that the three novice participants had insufficient previous financial knowledge to make this evaluation.

Research Question 5: Do semantic models provide useful knowledge to data modellers performing tasks related to a financial regulatory report?

Evaluation: The data modellers found the knowledge in the ontology and concept maps both effective and useful when answering the experimental questions. As the experiment did not measure the comprehension of the knowledge subsequent to the experimental sessions, it was not determined if the participants retained a useful understanding of the concepts.

6.7.2 *Was the model more useful for business or information knowledge?*

The participants rated the experimental models as helpful or very helpful in 88% of questions requiring business knowledge and in 62% of questions requiring information structure knowledge.

Both types and combinations of models were ranked as equally highly useful when answering business domain questions, which is supported by the observation that the participants frequently made use of the hyperlinks between the concept maps and ontology when answering these questions. Both Novice 2 and Novice 3 followed a similar pattern when answering business questions – firstly the concept map would be used as a navigation tool and then the hyperlink would be used to navigate to the ontology for definitions.

The data modellers were observed to show a preference for using the ontology when asked to provide an exhaustive list of sub-types or when the specification of a calculation. However, the three questions that were most frequently answered incorrectly or not attempted were those that required the participants to read the ontology to understand a subtle difference in the classification used in a report or to find the data model attributes required for a complex calculation. These questions were pre-classified as either medium or hard difficulty questions, so it is possible that

high numbers of incorrect answers due their difficulty rather than the type of knowledge being represented.

Research Question 6: Do semantic models better represent knowledge about the business domain or the structure of the information.

Evaluation: Business domain knowledge can be usefully represented in either technical models such as ontologies or in less formal models such as concept maps and structured vocabularies. While an ontology was well suited for precisely representing information structures, some complex representations resulted in knowledge that was difficult for data modellers to use.

6.7.3 Is a concept map or ontology model most useful in the repository?

While the ontology was identified as the single most useful models, it was also frequently found to be useful in combination with the concept maps. The concept maps were found to be helpful for getting a grasp of the structure of the concepts, were easy to understand and were especially useful for less specific conceptual questions. The ontology was good for answering questions that required specific definitions or calculation examples. The results suggest that the correctness of the answers and the helpful rating were independent of the preferred model for any given questions

The experimental observations suggest that the comparable usefulness of the models is influenced by an individual's preference for searching for knowledge either by visually scanning content or by using text based search tools. Once Expert 1 and Novice 2 successfully used the ontology tool search functionality to find definition for the first question, they typically started each subsequent question by searching the ontology. Novice 1 spend over five minute exploring the concept map at the start of the session and went on to use the concept map by itself or in combination with the ontology when answering all other questions.

The navigation functionally provided by the model tools influenced the usability of each of the models. The hyperlink links created between related concepts in the ontology and the concept maps was identified as an important usability feature by all three of the novice modellers. It should be noted that this feature was added almost as an afterthought in the concept map implementation described in Section 5.4. One

participant suggested that it should be possible to navigate from an ontology entry to the relevant part of the concept map. Another participant suggested that the description of calculations in the ontology would have been easier to understand if the calculation pseudo-code used hypertext links to other ontology entries.

Research Question 7: Is a concept map or an ontology more useful in the knowledge repository?

Evaluation: Individual preference is a determining factor of which type of model is most useful to any given data modellers. The combination of ontology and concept maps in the knowledge repository accommodated a variety of learning and content navigate styles.

6.7.4 Was the knowledge repository more useful to novice or expert users?

The mean of novice participant's responses to 'Please rate how helpful this model was in answering the question' across all questions was 4.1 compared to a mean of 3.5 for the experts. The analysis of the non-attempts, correct and incorrect answers in Figure 6.4 shows a similar distribution between both expert and novice participants. These two results suggest that the novices found the repository slightly more useful and that their low level of banking regulatory knowledge was not a barrier to the effective use of the semantic models to complete the experimental questions. It was observed the novice users typically spend more time both reading the questions and checking the answers they provided. The experts were more likely to rely on their own existing knowledge and then provide an answer without double checking with the content of the models.

A number of the participants did spend quite some time critiquing the content both in terms of how it differs from their understanding of semantic models and how they would have implemented the model differently. This is an example of the challenges of getting acceptance of semantic models from technical modellers that was identified by the subject matter expert interviews and described in Section 4.5.4. Both expert participants expressed the view that they were unsure of which model they were meant to use of each question, while in comparison the novices just gone on with the task at hand.

Vieritz *et al.* (2013) evaluation of a learning environment that included a knowledge model found that students strongly indicated that they would use a similar learning environment again. In comparison, while teachers would recommend that students use the learning environment, they would be less likely to use the environment for their own learning. That study's results contrasts strongly with the experimental feedback provides by both novice and expert modellers who all indicated that they would both use a similar knowledge repository again and recommend it to other data modellers.

Research Question 8: Are the semantic models more useful to novice or expert data modellers?

Evaluation: Both novice and experts found the models useful and all users would recommend the repository as a learning tool to novice modellers. The novice's use of the knowledge repository was as effective as the experts given with the novices answering questions with the similar level of correct answers as the experts.

6.8 Conclusions

The chapter presented the results and evaluation of the experiment that was successfully carried out using the approach and questions described in Section 6.2 and 6.3. An overview of the experimental session and the individuals who participated in the study was provided in Section 6.4. The participants successfully used both the ontology and concepts map to answer question in relation to the sample business reports and data model.

Sections 6.5 and 6.6 and presented a summary and analysis of the results of the ten experimental questions. The knowledge repository was rated as helpful or very helpful in 74% of all attempted question. The ontology was found to the most useful model overall especially when answering the information structure questions, but the combinations of the concept maps and ontology were useful when answering business domain questions. The results identified that individuals show a strong personal preference for using either the ontology or the concept map. All participants indicated that they would recommend a similar knowledge repository to other financial data modellers. Participants found the concept map accessible and highlighted the ease of use of the hyper-link integration between the concept map and the ontology.

The evaluation in Section 6.7 provided answers to the research questions related to the use of semantic models as a knowledge repository. The evaluation shows that the knowledge repository was at least as useful those explored in the studies of Vieritz *et al.* (2013) and Osman *et al.* (2011). It also identified the strong influence that both individual preference and technical implementation can have on the data modellers ability to understand and use the knowledge in a repository.

The discussion on data modeller knowledge acquisition in Chapter 4 identified the expectation that data modellers must acquire business domain knowledge but that semantic models were not generally seen as playing a role in such learning. The result in this chapter have demonstrated that such knowledge can be represented in a semantic models based repository and can be effectively used by data modellers when addressing low to medium complexity tasks. Data modellers should be seen as both creators and consumers of ontologies and other semantic models. This suggests that the emphasis that Moody and Shank's (2003) place on involving business users in the evaluation of data models could be similarly applied to the involvement in data modellers in evaluation of semantic models.

7 CONCLUSIONS AND FUTURE WORK

7.1 Introduction

This final chapter summarises the findings of the research conducted for this dissertation. It evaluates the approach taken, identifies the limitation of the research and proposes future work. Section 7.2 provides a summary of the research undertaken, the research objectives and how they were met by each of the dissertation chapters. Section 7.3 describes the contribution to the body of knowledge made by the research. Section 7.4 provides a critical evolution of the independent research, the implementation of the knowledge repository and the experimental design. Finally, Section 7.5 presents potential future research and work identified during this research project.

7.2 Research Definition & Research Overview

Data modellers in the banking industry are expected to utilise both technical and business knowledge when they model financial data required to meet both business and regulatory reporting requirements. The research explored and evaluated the combined use of an ontology and concept maps by data modellers when acquiring knowledge of financial and regulatory concepts.

Chapter 1 introduced an approach that divided the research activities into the three stages of background research, implementation and experimentation. The project met each of the research objectives that were presented in Section 1.5.

- A review of existing academic and industry literature was conducted to identify the use of ontologies, concept map and other types of semantic models in the financial industry.
- Interviews were conducted with two financial data modelling experts to understand the benefits and challenges in the use of the semantic models in the financial industry. The interviews also explored how data modellers acquire both business and technical knowledge.
- A knowledge repository comprising of an ontology and concept maps was implemented which allowed for an evaluation of the modelling and knowledge management challenges encountered

- An experiment was designed and successfully executed to assess the relative merits and disadvantages of the semantic models used to represent the business and technical knowledge required by data modellers tasked with understanding and interpreting financial regulatory report.
- The quantitative and qualitative results and finding of the experiment were documented and evaluated in this dissertation document.
- The results of the experiment were contrasted with the current views in the literature and suggestions were made for how both ontologies and concepts maps could be better used for financial information modelling.
- Recommendations for any future research in this area are made in Section 7.5.

Chapters 2 and 3 explored the topics of semantic models and relationship between semantic modelling and knowledge management. The literature informed the other research activities using the knowledge management themes of people, process and technology.

- *People:* Venable (1996) described how people become expert data modellers through the acquisition of specific knowledge. The challenges of knowledge acquisition faced by individual data modellers was explored during the interviews with experts in Chapter 4 and in the discussions on the experiment participants' use of knowledge repository in Chapter 6.
- *Process:* Abi-Lahoud *et al.* (2014) describe a semantic modelling process that performs a translation of banking knowledge into explicit model formats requiring inputs from both business and technical experts. The similarity between knowledge management processes and semantic modelling was discussed in Chapter 3 and evaluated in the model implementation described in Chapter 5.
- *Technology:* Both Bennet (2014) and the SME's interviewed in Chapter 4 suggested that technology focused semantic model implementations are likely to fail unless they also ensure that business benefits are realised and the model created is implementable. The data models quality factors suggested by Moody and Shanks (2003) were found to be a useful approach when evaluating the results of the experiment described in Chapter 6, especially when contrasting the usefulness of tooling functionality with the use of model content.

Chapter 4 discussed how the use of semantic models in the financial industry is influenced by changes in regulatory reporting requirements, evolutions in semantic modelling formats and numerous financial crises. The main types of models used in the industry include ontologies such as FIBO, and structured vocabularies such as SBVR. The emphasis is on the use of semantic models for interchange of information between IT systems or to facilitate clear communication between technical modellers and business users. There is little evidence that semantic models are used as knowledge repositories for technical experts such as data modellers to acquire business domain knowledge.

Chapter 5 discussed how an implementation of a semantic model based repository required a combination of the technical and business domain expertise. The use of two semantic model types encouraged the adoption of iterative implementation of knowledge content. Both semantic and knowledge modelling require modellers that follow Smith's (2003) Otologist's Credo to ensure they have understanding of the knowledge that they are representing. The financial industry trends for ontologies that are both formally correct and understandable by both business users and technical models suggest that the knowledge acquisition processes familiar to the knowledge management practitioners should be applied to semantic modelling.

Chapter 6 presented results that showed that data modellers found the knowledge in the ontology and concept maps both effective and useful when answering the experimental questions. The combination of ontology and concept maps accommodated a variety of learning styles and preferences observed as the participants used the knowledge repository. Business domain knowledge was usefully represented in both the ontologies and concept map, while the ontology was most suited for precise definitions and representing information structures.

7.3 Contributions to the Body of Knowledge

This research describe in this dissertation makes a number of contributions to the subject of the application of the combined use of concept maps and ontologies as part of knowledge management in the financial industry.

The literature and background research identified that while ontologies and structured languages are widely discussed as tools for solving the translation problem between business and IT experts, there is not a similar awareness of concept maps beyond their

use in financial education. This research shows that in the financial industry, the primarily objective of the use of semantic models is to ensure the correct and accurate implementation of business information and rules in IT systems, with a secondary objective of providing a common language of communication between humans. Even though significant effort is invested into translating business knowledge into semantic models, they are not viewed as part of the knowledge management landscape made up of content management repositories, knowledge portals and communities of practice. This research has described that the emphasis on semantic models as a technical artefact rather than a knowledge repository does not encourage their use by data modellers when acquiring business domain knowledge.

The evaluation of the implementation and use of the knowledge repository demonstrated a number of benefits of using a combination of both ontology and concept maps. The availability of the two model types supported a variety of personal user preferences for searching, browsing and making relationships between the concepts modelled in the repository. The implementation showed that the modeller can take advantage of the best aspects of the two model types when representing different types of knowledge - for example using the ontology for defining risk calculations and concepts map for representing risk policy conceptual cross-links.

The results of the experiment strongly suggest that the knowledge repository was effectively used by both novice and expert financial data modellers to answer questions with a range difficulty. While users found the repository effective when accessing business domain knowledge, they did encounter difficulties when access the most complex information structure knowledge. The results showed that the usefulness of the repository was strongly linked to the technical implementation of the hyper-links between the concept map and the ontology.

The research suggests that a semantic model based repository should comprise of two model components. A visual and non-intimidating format such as concept maps that provides an introduction or conceptual hook for users who are not familiar with the knowledge topic. The same concepts should then be represented in a formal model such as an ontology that both defined lower level concepts and provides specifications for classifications and calculations.

7.4 Experimentation, Evaluation and Limitation

The interviews and the experiment conducted as part of this research involved a relatively small number of participating experts and data modellers. While the experts interviewed both had a wide and varied experience using models in the financial industry, it would have been beneficial to confirm their opinions with similarly experienced individuals from the financial, regulatory and technology sectors. The data modellers who participated in the experiment had a variety of financial data modelling expertise which supported the design for the experiment and allowed for some interesting observations on the comparative usefulness of the models. A larger group of participating data modellers would be required to allow statistical conclusions to be made with regard to the experimental results.

The semantic model implementation was limited to an ontology and concept maps, each of which were deployed in a single tooling option. An extension of the implementation to include an SBRL and OWL implementation of the knowledge would have allowed for a direct comparison with the findings of Abi-Lahoud *et al.* (2014). The implemented model content was developed without independent review by a Basel regulation subject matter expert, as the individual planned for this activity was not available at the time of implementation. The lack of expert review of the model content introduced the risk that the financial business domain knowledge represented in the model contained errors or omissions. This risk was mitigated somewhat by the fact that some of the experimental participants are expert financial data modellers and their feedback did not identify a serious model content issues.

The choice of having a hands-on experimental approach worked in practice as it ensured that the data modellers were engaged with the implemented model. This allowed them to give feedback based on their experience using the knowledge repository, rather than providing a passive critique of the model content. Asking the data modellers to perform a series of increasingly difficult tasks in a limited time provoked them into giving honest feedback on the benefits and limitation of the different models. However as the experiment did not measure the comprehension of the knowledge at any time the after experimental session, it was not determined if the knowledge was internalised by the participants. A limitation of this experimental approach is that it would require significant time and face-to-face access to a large population of the data models to expand the size of the study group.

7.5 Future Work and Research

The research described in this dissertation suggests future work that could either extend the scope of the research or address some of the limitations of this research. The following research suggestions could be applied to domains other than the financial industry and could be aligned to similar studies in other regulated domains such as medicine and the health sciences.

A broader understanding of the benefits and challenges of using semantic models in the financial industry could be gathered through a combination of surveys and interviews. The sample group should be selected representative of both business and technical subject matter experts. The technical experts should include data modellers, semantic modellers, knowledge modellers, database administrator, report developers and data integration specialists. The business experts should include a mixture of regulators, academic, data governance officers, members of the FIBO initiative and line of business users.

The scope of the knowledge modelling described in Chapter 5 was relatively small and was conducted by an individual modeller. Future research could examine if the benefits of an iterative development using the two semantic model types would also be observed in a collaborative modelling situation with a team of modellers. This research could explore the similarities between the processes of semantic modelling and knowledge acquisition in an environment that provided multiple formats for knowledge modelling.

The use of algorithms to transform a semantic model to a data model was discussed in Chapter 2 and Chapter 4. Research could be conducted to compare the quality of data models derived either algorithmically from a semantic model or created manually by a data modeller who has access to the same semantic model. The data model quality factors suggested by (Moody and Shanks, 2003) could be used as a framework for comparison.

The knowledge repository described research could be evaluated over a longer time scale by making it made available to a group of data modellers as they perform their day to day modelling activities. This could evaluate the long term acquisition of the knowledge by the data modellers by testing their understanding when they did not have direct access to the repository.

BIBLIOGRAPHY

Abi-Lahoud, E, Butler, T and O'Brien, L, (2014), 'A Hermeneutic Approach To Solving The Translation Problem In Designing Ontologies', in *22nd European Conference on Information Systems, ECIS 2014 Conference*.

Ackoff, R (1989), From data to wisdom, *Journal of Applied Systems Analysis*, Vol 16.

Åhlberg, M K (2013), Concept mapping as an empowering method to promote learning, thinking, teaching and research, *Journal for Educators, Teachers and Trainers*, Vol 4 (1) , pp. 25–35.

AIB Group (2013), *AIB Pillar 3 Disclosures 2013*. Available:

<http://investorrelations.aib.ie/content/dam/aib/investorrelations/docs/resultscentre/pillar3/pillar-3-2013.pdf>, [Accessed 18th Dec 2014]

Albarrak, K M and Sibley, E H, (2009) 'Translating Relational & Object-Relational Database models into OWL models', in *IEEE International Conference on Information Reuse Integration, 2009. IRI 2009*, pp. 336–341.

Angles, R (2012). 'A Comparison of Current Graph Database Models', in *28th International Conference on Data Engineering Workshops, IEEE ICDEW 2012*, pp. 171–177.

Basel Committee on Banking Supervision (2013). *Principles for Effective Risk Data Aggregation*. Available: <http://www.bis.org/publ/bcbs128.pdf>, [Accessed 4th September 2014]

Basel Committee on Banking Supervision (2006), *International Convergence of Capital Measurement and Capital Standards*. Available: <http://www.bis.org/publ/bcbs239.pdf>, [Accessed 18th December 2014]

Bergman, M (2007) *An Intrepid Guide to Ontologies*. Available:

<http://www.mkbergman.com/374/an-intrepid-guide-to-ontologies/>, [Accessed 18th December 2014].

Bhatt, G D (2001), Knowledge management in organizations: examining the interaction between technologies, techniques, and people, *Journal of Knowledge Management*, Vol 5 (1), pp. 68–75.

- Blaha, M (2010), *Patterns of Data Modeling*, CRC Press, Boca Raton, FL.
- Blair, D C (2002), Knowledge Management: Hype, Hope, or Help?, *Journal of the American Society for Information Science & Technology*, Vol 53 (12), pp. 1019–28.
- Bonson-Ponte, E, Escobar-Rodriguez, T and Flores-Munoz, F (2007), The role of metadata language implementation in the European banking supervision network, *International Journal of Networking and Virtual Organisations*, Vol 4 (3), pp. 245–56.
- Chisholm, M D (2010), *Definitions in Information Management*, byDesign Media, Canada.
- Correndo, G and Alani, H, (2007), ‘Survey of Tools for Collaborative Knowledge Construction and Sharing’ in *2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology Workshops*, pp. 7–10.
- De Vergara, J E L and Villagra, VA, Berrocal, J (2004), Applying the Web ontology language to management information definitions, *IEEE Communications Magazine*, Vol. 42 (7), pp. 68–74.
- EDM Council (2014), *Financial Industry Business Ontology*. Available: <http://www.edmcouncil.org/financialbusiness> [Accessed 12th December 2014]
- Fensel, D (2003), *Ontologies: a silver bullet for knowledge management and electronic commerce*, 2nd edn, rev. and expanded edn, Springer-Verlag, Berlin.
- Graudina, V and Grundspenkis, J, (2011), ‘Algorithm of Concept Map Transformation to Ontology for Usage in Intelligent Knowledge Assessment System’, in *Proceedings of the 12th International Conference on Computer Systems and Technologies*, CompSysTech 2011, New York, NY, USA, pp. 109–114.
- Greenberg, R K and Wilner, N A (2015), Using concept maps to provide an integrative framework for teaching the cost or managerial accounting course, *Journal of Accounting Education*, Vol. 33 (1), pp. 16–35.
- Gurteen, D (1999), Creating a knowledge sharing culture, *Knowledge Management Magazine*, Vol. 2 (5).
- Hayes, P, Eskridge, T C, Saavedra, R, Reichherzer, T, Mehrotra, M and Bobrovnikoff, D (2005), ‘Collaborative knowledge capture in ontologies’, in *Proceedings of the 3rd International Conference on Knowledge Capture*, pp. 99–106.

Kimball, R (2011), *The Evolving Role of the Enterprise Data Warehouse in the Era of Big Data Analytics*. Available:

<http://ccsglobaltech.com/pdf/DataWarehouseBigDataAnalyticsKimball.pdf> [Accessed 7th November 2014]

Kimball, R and Ross, M (2011) *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, John Wiley & Sons.

Kramar, V (2013), 'Data Models for Home Services', in *Proceeding of the 13th Conference of FRUCT (Finnish-Russian University Cooperation in Telecommunications) Association*, pp. 72–84.

Kuper, G M and Vardi, M Y (1993), The Logical Data Model, *ACM Transactions Database Systems*, Vol 18 (3), pp. 379–413.

Malik, P (2013), Governing Big Data: Principles and practices, *IBM Journal of Research and Development*, Vol. 57, p. 1.

Milton, N R (2007), *Knowledge Acquisition in Practice: A Step-by-step Guide*, Springer Science & Business Media, London.

Moody, D L and Shanks, G G (2003), Improving the quality of data models: empirical validation of a quality management framework, *Information Systems*, Vol. 28 (6), pp. 619–650.

Nesbit, J C and Adesope, O O (2006), Learning With Concept and Knowledge Maps: A Meta-Analysis, *Review of Educational Research*, Vol. 76, pp. 413–448.

Nonaka, I, Toyama, R and Konno, N (2000). *SECI, Ba and Leadership: a Unified Model of Dynamic Knowledge Creation*, *Long Range Planning* Vol. 33 (1) , pp. 5–34.

Novak, J D and Cañas, A J (2006), *The Theory Underlying Concept Maps and How to Construct Them*, Florida Institute for Human and Machine Cognition. Available:

<http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf> [Accessed: 5th May 2014]

Noy, N F and McGuinness, D L (2001), *Ontology development 101: A guide to creating your first ontology*. Available:

http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html [Accessed: 19th February 2015]

Object Management Group (2015), *Semantics of Business Vocabulary and Business Rules Specification Document Version 1.3*. Available:

<http://www.omg.org/spec/SBVR/1.3/PDF> [Accessed 13th June 2105]

Obrst, L (2003), 'Ontologies for semantically interoperable systems' in *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, ACM, pp. 366–369.

Obrst, L (2008), *The Ontology Spectrum & Semantic Models*. Available:

<http://semanticcommunity.info/@api/deki/files/5954/leobr05122008.pdf> [Accessed 14th February 2015]

Osman, N A, Ho, S B and Haw, S C (2011), 'Empirical approach in ontology engineering: An assessment on novice modelers', in *Software Engineering (MySEC), 2011 5th Malaysian Conference*, pp. 501–507.

O'Sullivan, P, Thompson, G and Clifford, A (2014) Applying data models to big data architectures, *IBM Journal of Research and Development*, Vol 58 (5/6).

Pope, A and Butler, T (2012), 'Unpacking the People, Process and Technology Dimensions of organisational KMS', in *ECIS 2012 Proceedings*, European Conference on Information Systems (ECIS), paper 183.

Reserve Bank of New Zealand (2007) *Capital adequacy ratios for banks - simplified explanation and example of calculation*. Available:

http://people.stern.nyu.edu/igiddy/articles/capital_adequacy_calculation.pdf [Accessed 9th April 2015]

Rodriguez-Priego, E, García-Izquierdo, F J and Rubio, Á L (2010), 'Modeling issues: a survival guide for a non-expert modeler' in *Model Driven Engineering Languages and Systems*. Springer, pp. 361–375.

Rowley, J (2007), The wisdom hierarchy: representations of the DIKW hierarchy, *Journal of Information Science*, Vol. 33 (2), pp. 163–180.

Sadalage, P J and Fowler, M (2013), *NoSQL distilled: a brief guide to the emerging world of polyglot persistence*, Addison-Wesley, Upper Saddle River, NJ.

Simon, J (2007), Concept Mapping in a Financial Accounting Theory Course, *Accounting Education*, Vol. 16 (3), pp. 273–308.

- Smith, B (2003), 'Ontology' in L Floridi (ed), *Blackwell Guide to the Philosophy of Computing and Information*, Wiley-Blackwell, pp. 155–166.
- Smith, B, Kusnierczyk, W, Schober, D and Ceusters, W (2006), 'Towards a Reference Terminology for Ontology Research and Development in the Biomedical Domain', in *Proceedings of KR-MED 2006*, Second International Workshop on Formal Biomedical Knowledge Representation, pp. 57–66.
- Venable, J R (1996), 'Teaching novice conceptual data modellers to become experts', in *Software Engineering: Education and Practice, 1996. Proceedings. International Conference*, pp. 50–56.
- Vieritz, H, Schmitz, H-C, Law, E L-C, Scheffel, M, Schilberg, D and Jeschke, S (2013), 'A Knowledge Map Tool for Supporting Learning in Information Science', in *Proceedings of the International Conference on Computer Supported Education (CSEDU)*, pp. 717–723.
- W3C, 2012. *OWL 2 Web Ontology Language Primer (Second Edition)*. Available: <http://www.w3.org/TR/2012/REC-owl2-primer-20121211/> [Accessed 21st February 2015]
- W3C (2014), *RDF 1.1 Primer*, Available: <http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/> [Accessed 21st February 2015]
- W3C (2015), *W3C Semantic Web*. Available: <http://www.w3.org/standards/semanticweb/> [Accessed 19th February 2015]
- Wagner, C (2006), Breaking the knowledge acquisition bottleneck through conversational knowledge management, *Information Resources Management Journal*, Vol. 19 (1), 70–83.
- Zachman, J A (1987), A framework for information systems architecture, *IBM Systems Journal*, Vol. 26 (3), pp. 276–292.
- Zhu, J and Huang, Z (2014), Banks' Micro Enterprises Loan Credit Risk Decision-making Model Innovation in the Era of Big Data and Internet Finance, *Journal of Management and Strategy*, Vol. 5 (2) , pp. 63-69

A. SME INTERVIEW QUESTIONS

Data modelling and industry business knowledge

- How long have you been working with data models?
- Which industries or business sectors have you developed data models for?
- Would you describe yourself as a data modeller, business analyst or information architect
 - How would you compare these roles?
- Were there differences between how you have gained technical data modelling knowledge and industry business knowledge?

Experience and View using Semantic Models

- How would you define a semantic model?
 - What types of semantic models have you worked with?
 - How does it differ from a data model?
- What is your experience working with ontologies?
 - Do you think that ontologies are technical models or business?
 - What are the challenges have you faced when creating ontologies?
- How are semantical models used in conjunction with relational data models?
 - What are tasks or types of users are they most suited to?

Semantic in the Financial Industry

- What are the significant trends in the use of semantic models used in the Financial Industry?
- What need is being met by financial institutions ontologies
 - Regulatory reporting standardisation
 - Information Exchange
 - Data Governance such as Risk Data Aggregation
- In your opinion, how effective have ontologies been in the financial industry?
 - What are the challenges of their adoption
 - Have they fulfilled their promise?

Advice for novice financial data modellers

- How would you suggest a data modeller could best learn about financial concepts or data modelling techniques?
- Looking back on how you learned about financial concepts, what worked well for you?
 - Are there any approaches that did not work so well?
- Do you think that industry standard ontologies such as FIBO are useful for novice learning about financial concepts?

B. INTERVIEW KNOWLEDGE ANALYSIS

A listing of the concept categories, concepts and the count of the number of times when concept code was used in the transcript is provided in Table B.1. The coded transcripts are provided in the supporting material.

Table B.1 : Occurrences of Concepts in Interview

Concept Category	Count of Codes in Transcripts
Actors in Semantic Modelling	
Business - Business User, Regulator, Industry Body	22
Modeller – Data Modeller, Business Analyst, Information Architect	17
IT – IT Systems, Modelling Tools	8
Data Modeller Knowledge Acquisition	
Modelling Experience – Industry Experience, Expert Modeller Behaviour	16
Informal Learning – Learning by doing, reading, conversations	13
Formal Training	7
Semantic Models	
Ontology	17
Semantic Models (Non-Specific)	14
Conceptual Models – Structured Languages, Classification Models	11
Data Model	4
Semantic Model Challenges	
Limited Acceptance of Model	18
Hard to Implement	8
Difficult to Represent Knowledge	6
Bridging Business and IT Gap	4

C. TRIADIC AND CARD SORT FOR SMES

Concepts Selection for Knowledge Elicitation

The concepts used in the triadic and card sort are presented in Figure C.1. Note the concept of ‘Concept Maps’ was replaced with ‘Conceptual Model’ as the participant were not familiar with C-Maps.

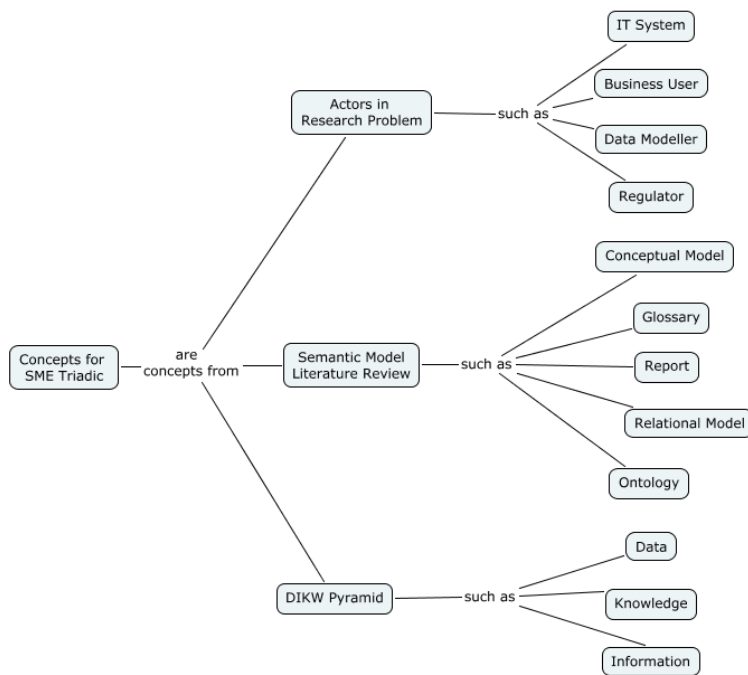


Figure C.1 : Concepts selected for Card Sort and Triadic

Card Sort Results

The first card sort performed by each of the SME's resulted in the categorisation of the concepts shown Figure C.2. Both categorisations identified the user roles and emphasises that semantics models are design objects that are used in the development of IT end products.

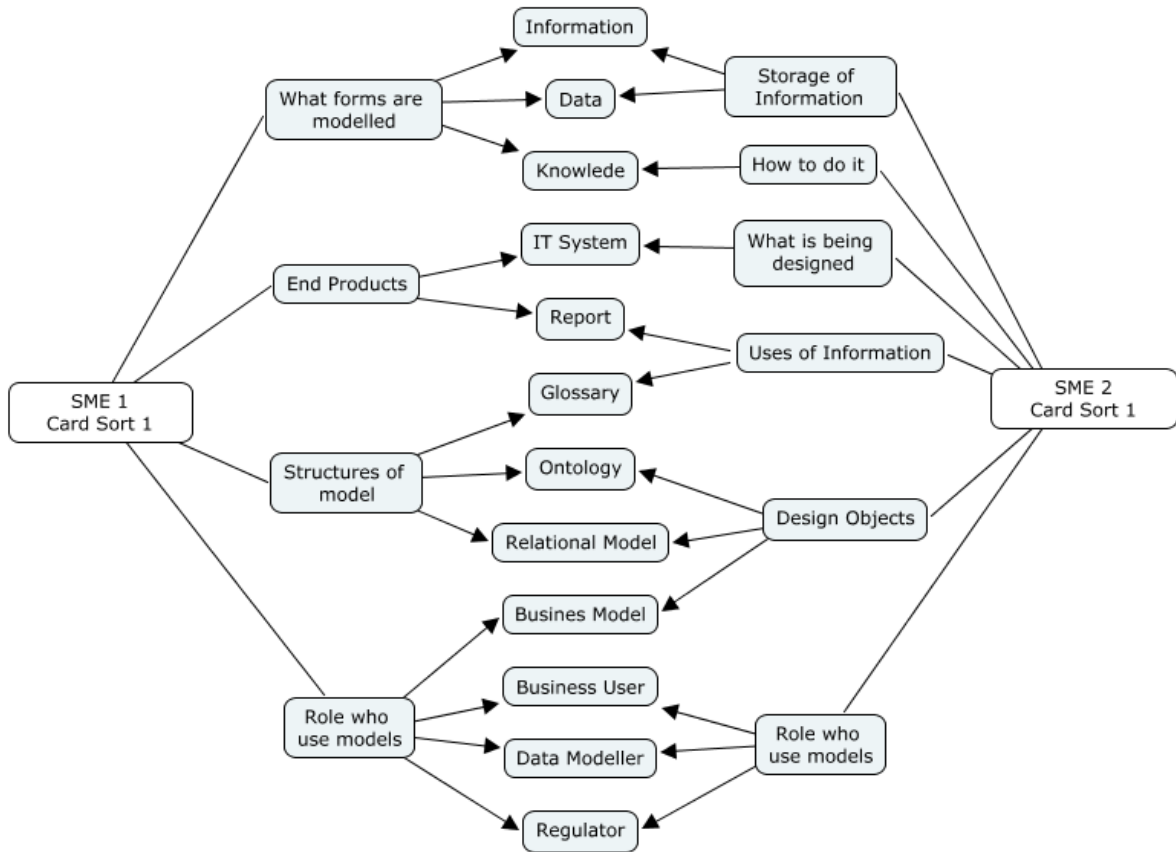


Figure C.2 : Results of First Card Sort by Subject Matter Experts

The second card sort performed by the SMEs resulted in the three comparable but overlapping groupings shown in Figure C.3. Semantic models such as are used by non-technical individuals who need to produce or access business and regulator knowledge. These models require a technical infrastructure to create or relate the knowledge to data and information. There is middle ground inhabited by data modellers who create semantic models. While both SMEs identify that the data modeller sits between the technical and non-technical world, there is a difference of opinion on what types of semantic models fall into this category.

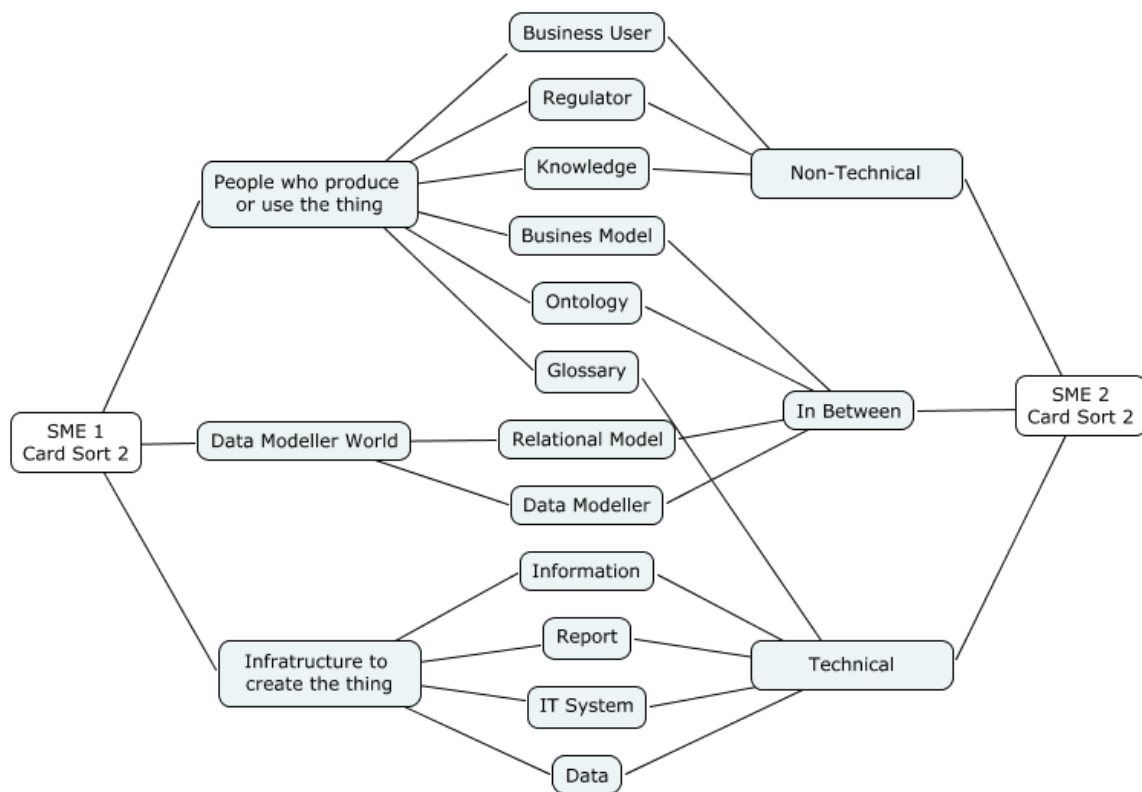


Figure C.3 : Results of Second Card Sort by Subject Matter Experts.

Triadic Results

Each subject matter expert was asked to perform four triadic comparison with concepts selected randomly from the 12 concepts prepared for the card sort. Three cards were dealt and the interviewee was asked to identify a way in which two of the concepts of them are similar but different from the other concept.

The results of the triadic comparison are provided in Table C.1.

Table C.1 : Triadic Results

Individual	Triadic	Concept Pair	Third Concept
SME 1	1	Data is presented in a Report	Glossary is metadata that tells you the meaning of data.
SME 1	2	Data Modeller determines the structure s of an IT System	Relational Model express a specific implementation
SME 1	3	Ontology is an expression of concepts that makes sense to Business Users	Information is something that is contained.
SME 1	4	Business Models is used by Regulator to express what need to be reported	Knowledge of a regulator is encapsulated in a business model
SME 2	1	Business Model & Information are more nebulous or abstract	Glossary is an instantiation
SME 2	2	Report & Business User are linked by a business issue or business area or context	Data is nebulous or abstract
SME 2	3	Knowledge & Relational Model are not a role.	Regulator is a role
SME 2	4	Data Modeller & Ontology are design time assets or activity	IT System are, a run-time physical thing where we deploy

D. EXPERIMENT LOGICAL DATA MODEL

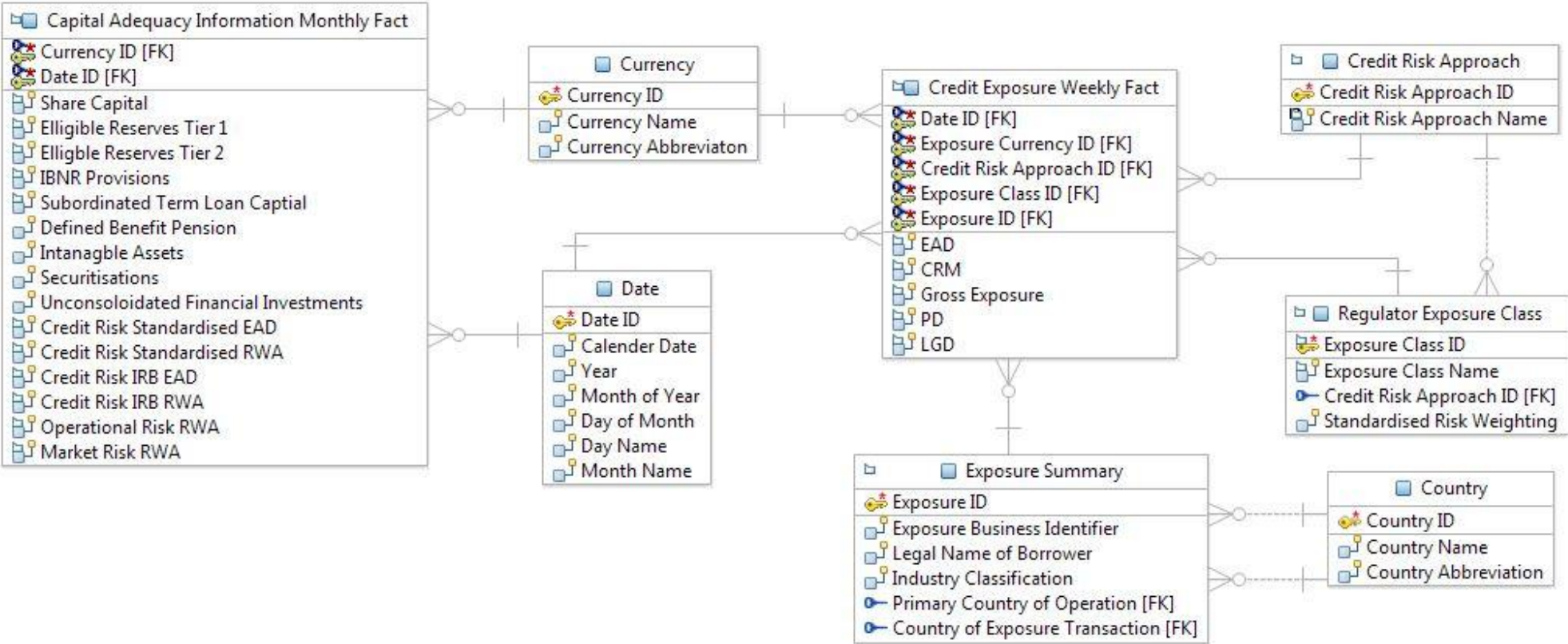


Figure D.1 : Logical Data Model from Experiment

E. EXPERIMENT CONCEPT MAPS

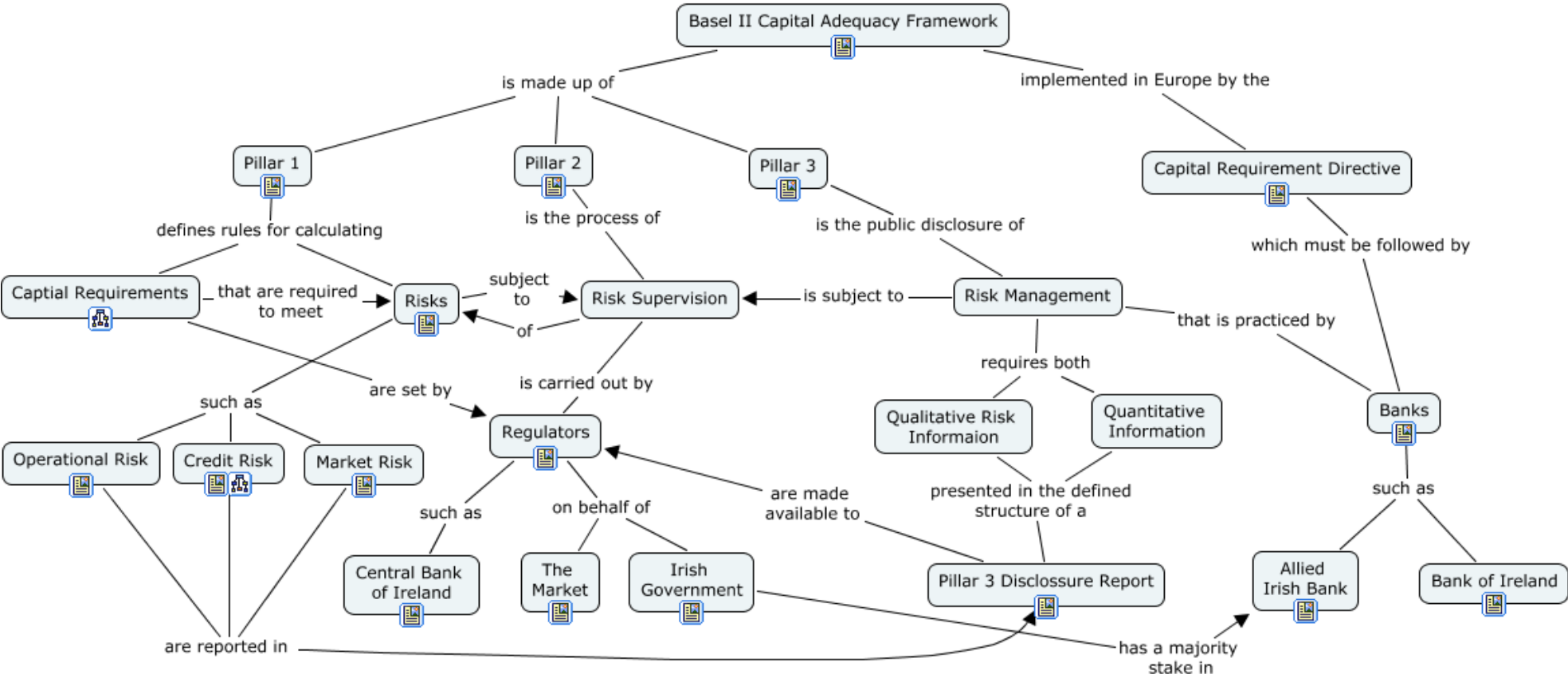


Figure E.1 : Basel II Capital Adequacy Framework Concept Map

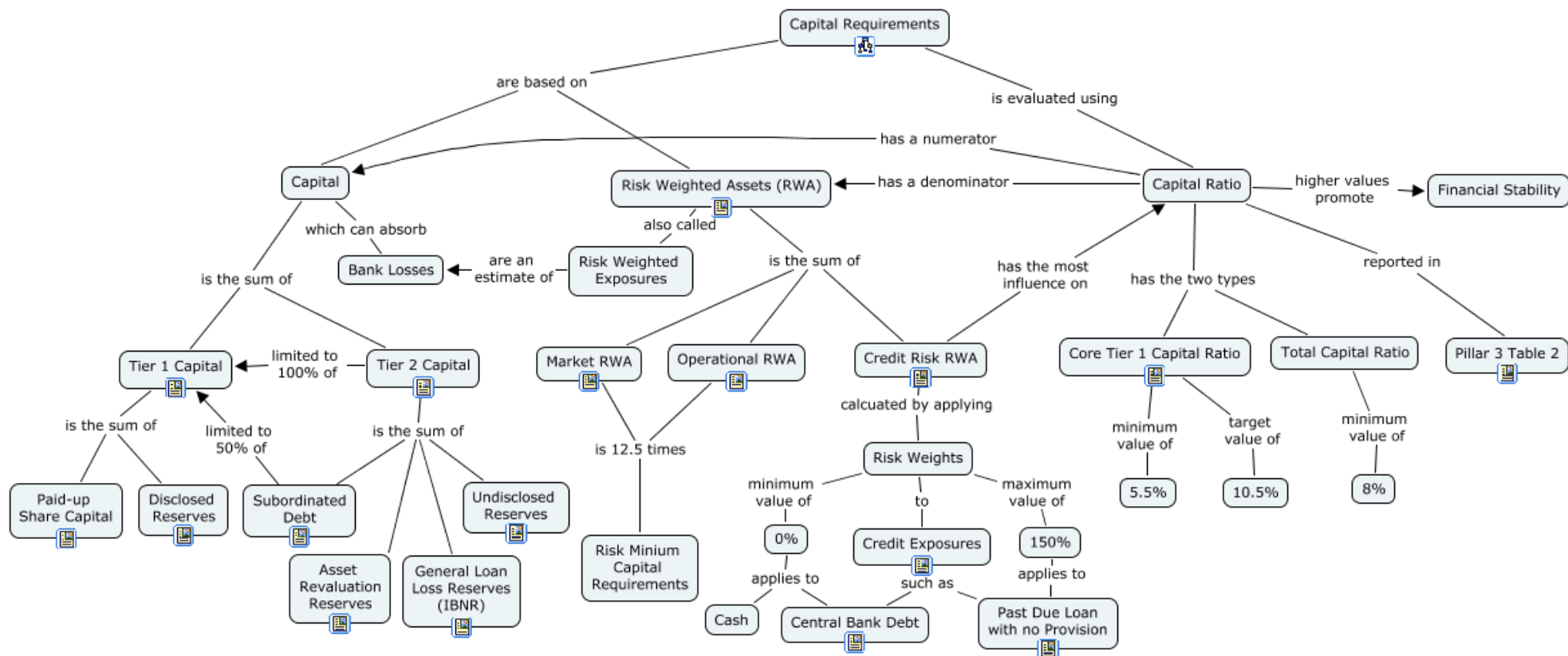


Figure E.2 : Credit Risk Concept Map

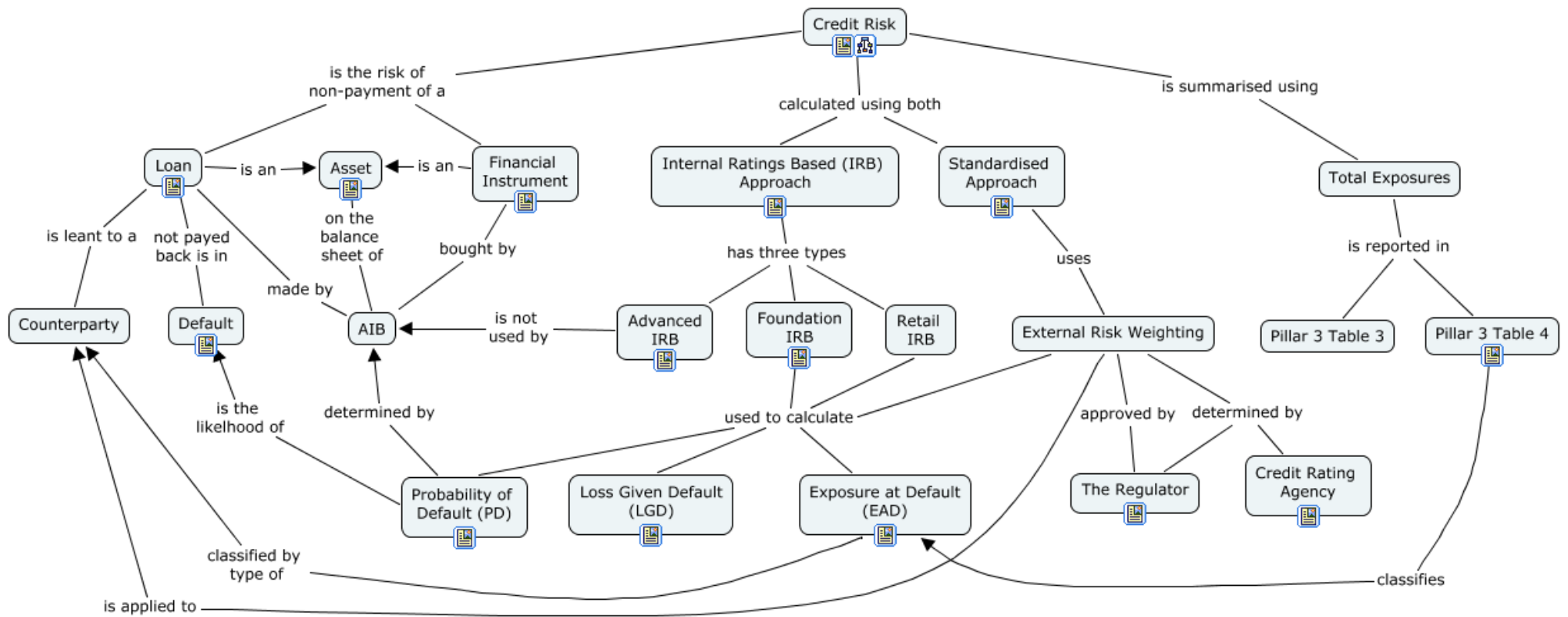


Figure E.3 : Capital Requirements Concept Map

F. EXPERIMENT ONTOLOGY TERM LIST

The table below provides the full list of the terms names contained in the ontology implemented in IGC. There is an export of the ontology terms in provided in IGG XML format as part the supporting material.

The table indicates where the term in the ontology had a corresponding entry in the;

- **Report**, where the term was used in a report field, column title or row title.
- **C-Map**, where the term has a corresponding concept in one of the three concept maps. In these cases there is a hyperlink from the concept to the ontology entry.
- **Data Model**, where the ontology term is assigned to a data model attribute or entity term list.

Table F.1 : List of Ontology Terms

Term Name	Ontology	Report	C-Map	Data Model
Administrative Bodies	Yes	Yes		
Advanced IRB	Yes		Yes	Yes
Allied Irish Bank	Yes		Yes	
Asset	Yes		Yes	
Asset Revaluation Reserves	Yes		Yes	
Bank	Yes		Yes	
Bank of Ireland	Yes		Yes	
Basel II	Yes		Yes	
Basel III	Yes			
Capital Requirements Directive	Yes		Yes	
Central Bank of Ireland	Yes		Yes	
Central Government and Central Banks	Yes	Yes		
Central Government and Central Banks (IRB)	Yes	Yes		
Core Tier 1 Capital	Yes	Yes	Yes	
Core Tier 1 Ratio	Yes	Yes	Yes	
Corporates	Yes	Yes		
Corporates (IRB)	Yes	Yes		
Counterparty Credit Exposure	Yes			
Covered Bonds	Yes	Yes		
Credit Conversion Factor	Yes			
Credit Rating Agency	Yes		Yes	
Credit Risk	Yes	Yes	Yes	Yes
Credit Risk Capital Requirement Approach	Yes			
Credit Risk Mitigation	Yes			Yes

Term Name	Ontology	Report	C-Map	Data Model
Credit RWA	Yes	Yes	Yes	
Default	Yes		Yes	
Disclosed Reserves	Yes	Yes	Yes	Yes
Expected Loss	Yes			
Exposure	Yes	Yes	Yes	
Exposure At Default	Yes	Yes	Yes	Yes
Exposure Weighted Average LGD	Yes			
Exposure Weighted Average Risk Weight	Yes			
Financial Instrument	Yes		Yes	
Foundation IRB	Yes		Yes	Yes
General Load and Loss Reserves	Yes	Yes	Yes	Yes
Gross Exposure	Yes			Yes
Institutions (IRB)	Yes	Yes		
Institutions (Standardised)	Yes	Yes		
Internal Ratings Based Approach	Yes	Yes	Yes	
IRB Exposure Classification	Yes	Yes		
Irish Government	Yes		Yes	
Items belonging to regulatory high risk categories	Yes	Yes		
Liquidity Risk	Yes			
Loan	Yes		Yes	
Loss Given Default	Yes		Yes	
Market Risk	Yes		Yes	
Market RWA	Yes	Yes	Yes	Yes
Operational risk	Yes		Yes	
Operational RWA	Yes	Yes	Yes	Yes
Organisation	Yes			
Other items	Yes	Yes		
Paid-Up Share Capital	Yes	Yes	Yes	Yes
Party	Yes			
Past due items	Yes	Yes		
Pillar 1	Yes		Yes	
Pillar 2	Yes		Yes	
Pillar 3	Yes		Yes	
Probability of Default	Yes		Yes	
Real Estate Collateral	Yes	Yes		
Regional Governments or Local Authorities	Yes	Yes		
Regulator	Yes		Yes	
Regulatory Capital	Yes		Yes	
Regulatory Retail	Yes	Yes		
Retail (IRB)	Yes	Yes		
Risk	Yes		Yes	
Risk Weighted Assets	Yes	Yes	Yes	Yes
Securitisation	Yes	Yes		
Securitisation Positions	Yes	Yes		

Term Name	Ontology	Report	C-Map	Data Model
Standardised Approach	Yes	Yes	Yes	Yes
Standardised Approach	Yes	Yes	Yes	Yes
Standardised Approach Exposure Classification	Yes	Yes		
Subordinated Debt	Yes	Yes	Yes	Yes
Table 2 - Capital Adequacy Information	Yes	Yes		
Table 3 - Group Capital Adequacy Information	Yes	Yes		
Table 4 - Total Exposures and Minimal Capital Requirements by Exposure Class	Yes	Yes		
Table 5 - Industry Distribution of EAD - Standardised Approach	Yes	Yes		
Table 6 - Geographic Distribution of EAD - Standardised Approach	Yes	Yes		
The Market	Yes		Yes	
Tier 1 Capital	Yes	Yes	Yes	
Tier 2 Capital	Yes	Yes	Yes	
Undisclosed Reserves	Yes	Yes	Yes	Yes

G. LIST OF SUPPORTING MATERIAL

The following supporting material is provided on the CD

Table G.1 : List of Supporting Material.

Name	Description	Format(s)
Concept Maps and Ontology Overview	Narrated video demonstration of the concepts maps and ontology described in Chapter 5.	MP4
SME1 Interview with Codes	Full transcript of interview with SME1 annotated with concept codes.	PDF
SME2 Interview with Codes	Full transcript of interview with SME2 annotated with concept codes.	PDF
Experiment Script and Answer Sheet	Introduction used by author when introducing the experiment to participants. Answers to the ten questions.	PDF
Experiment Questionnaire	Ten questions followed by the participants in the experimental study.	PDF
Experiment Result Dataset	Quantitative result from experiment including correctness, most useful model and helpful rating metrics	CSV
Experiment Observations	Qualitative notes and observations made by author during the experimental sessions.	PDF
Capital and Risk Datamart Logical Data Model	Logical data model derived to support the five sample reports and used as input for implementation and experiment. Requires IBM InfoSphere Data Architect to open. Image of model provided in Appendix D.	IDA LDM
Capital and Risk Ontology	Exports of terms from the experimental ontology content from IBM Information Governance Catalog. The XML format includes the references between objects in the ontology.	IGC XML IGC CSV Export
Capital and Risk Concept Maps	Three concept maps implemented in the experimental repository. Provided in both Cmap Tools and HTML format.	Cmap Tools HTML
AIB Pillar 3 Disclosure 2013	Regulatory report which is source of five sample reports used as input for implementation and experiment	PDF