



UNIVERSIDADE DE ÉVORA

ONTOLOGIAS COM SUPORTE EM METADADOS PARA INTEROPERABILIDADE ENTRE ARQUITETURA EMPRESARIAL E BUSINESS INTELLIGENCE

Francisco José Rosales Santana Guimarães

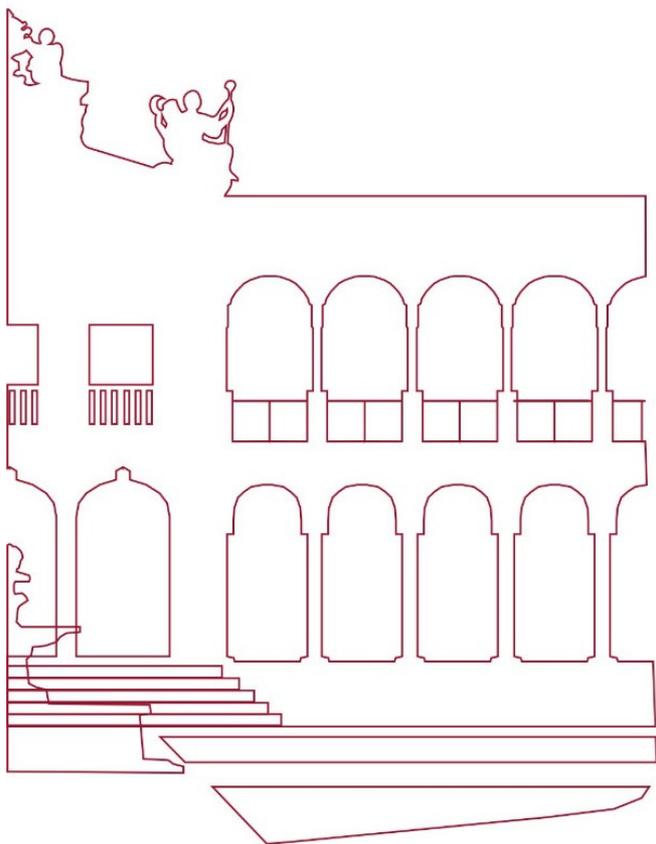
Orientadores: Professor Doutor Carlos Pampulim Caldeira

Professor Doutor Paulo Quaresma

Tese apresentada à Universidade de Évora para obtenção do
Grau de Doutor em Informática

Especialidade: Informática

Évora, 2018



INSTITUTO DE INVESTIGAÇÃO E FORMAÇÃO
AVANÇADA



ONTOLOGIAS COM SUPORTE EM METADADOS PARA INTEROPERABILIDADE ENTRE ARQUITETURA EMPRESARIAL E BUSINESS INTELLIGENCE

Francisco José Rosales Santana Guimarães

Orientadores: Professor Doutor Carlos Pampulim Caldeira
Professor Doutor Paulo Quaresma

Tese apresentada à Universidade de Évora para obtenção do
Grau de Doutor em Informática

Especialidade: Informática

Évora, 2018



Dedicado a:

Gabriela Peste

Gato Tricky

Emília Guimarães

Francisco Santana Guimarães

Carla Santana Guimarães

Aos meus amigos

Agradecimentos:

Aos meus orientadores por todo o apoio, ideias e soluções.

Ao Luís Estevens e Ricardo Sousa da EDIA.

Ao Tomás Salvado e Ricardo Lobão da Patris Fincor

À Noélia Peixoto e equipa de Suporte à Decisão da CGD



I. Resumo

Uma empresa é uma forma de organização que pode ser vista como um sistema, e como tal, passível de ser representada através de um modelo que permita capturar os conceitos que a definem considerando a sua estrutura (e.g. cliente, canal, produto) e semântica (e.g. relação entre cliente e canal). Com base neste modelo torna-se possível conhecer de forma holística a organização, delinear estratégias, implementar sistemas de informação e monitorizar o seu desempenho.

A modelação de uma organização ocorre em processos cíclicos, como é o caso de planeamento estratégico, desenho de processos ou desenho de soluções de sistemas de informação. Nestes processos, utilizam-se vários modelos que em conjunto representam o conhecimento sobre a própria organização. Apesar de existirem vários modelos, no contexto da nossa investigação consideramos que a arquitetura empresarial (AE) apresenta um referencial agregador e que permite criar uma visão holística da organização, além de que é suportada por linguagens de notação (e.g. UML, ArchiMate) e pode-se utilizar ferramentas para gestão destes modelos, o que possibilita uma normalização e reutilização de componentes. Para além da utilização numa perspetiva de gestão da organização, a AE pode ser igualmente a base para se detalhar requisitos funcionais e técnicos para implementação em aplicações informáticas no domínio de sistemas de informação. Estas aplicações podem ser operacionais, orientadas a tornar os processos mais eficientes e eficazes, ou informacionais, referidas nesta tese como sistemas de *business intelligence* (BI), focadas em monitorizar o desempenho através de eventos ocorridos (métricas), analisados em perspetivas de negócio (dimensões).

Cada modelo, independentemente da sua visão gráfica, é instanciado sob a forma de dados, mas com definições destes dados representadas em estruturas conhecidas como metadados. Como tal, os metadados são normalmente vistos como dados sobre dados. A importância dos metadados é hoje reconhecida no domínio da AE, *data governance* e BI. No entanto, um dos problemas está associado à ausência de interoperabilidade entre estes sistemas e seus metadados, visto que é necessária uma transposição entre modelos, desde a definição até ao suporte ao funcionamento da organização ao nível do seu sistema de informação. A falta de interoperabilidade, leva a um desalinhamento entre os conceitos utilizados em AE e a sua implementação em sistemas de informação, com impacto ao nível do esforço para a implementação e manutenção destes sistemas. Em particular, o desalinhamento apresenta maior impacto no caso dos sistemas de BI que utilizam conceitos de negócio sob a forma de dimensões e métricas que devem estar alinhados com os que são utilizados na definição da organização e dos seus objetivos estratégicos, desenhados na AE.

Esta tese centra a sua investigação neste problema de interoperabilidade no domínio da representação de conhecimento, considerando uma arquitetura de solução em que os metadados são vistos como uma ontologia organizacional, modelada em AE, enriquecida com outros conceitos (e.g. glossários, definição de bases de dados), reutilizada em sistemas de BI e permitindo a sua utilização em interfaces com utilizador baseada em processamento de língua natural. Reduz-se desta forma o esforço e complexidade de gestão de AE e BI, além de que torna assim possível um melhor alinhamento entre os conceitos utilizados na medição de desempenho e os conceitos definidos na modelação da organização, enquanto contributo para o alinhamento entre o negócio e sistemas de informação. Esta hipótese de solução define assim uma arquitetura e abordagem para suporte à inteligência organizacional.

Palavras-chave: Arquitetura empresarial, modelos de negócio, processos de negócio, *business intelligence*, metadados, *data lineage*, *data discovery*, *knowledge discovery*, *knowledge management*, *data governance*, ontologias, *ontology learning*, *reasoning*.



II. Abstract

Ontologies with Metadata Support for Interoperability between Business Architecture and Business Intelligence

A company is a form of organization that can be seen as a system, and as such, capable of being represented through a model that allows to capture the concepts that define it considering its structure (e.g. client, channel, product) and semantics (e.g. relationship between client and channel). Based on this model it becomes possible to know the organization in a holistic way, to delineate strategies, to implement information systems and to monitor its performance.

The modelling of an organization occurs in cyclical processes, as is the case of strategic planning, process design or design of information systems solutions. In these processes, several models are used, which together represent knowledge about the organization itself. Although there are several models, in the context of our research we consider that the enterprise architecture (EA) presents an aggregate reference and allows to create a holistic view of the organization, besides being supported by notation languages (e.g. UML, ArchiMate) it can use tools for the management of these models, which enables standardization and reuse of components. In addition to being used in a management perspective of the organization, the EA may also be the basis for detailing functional and technical requirements for implementation in IT applications in the field of information systems. These applications can be operational, oriented to make the processes more efficient and effective, or informational, referred to, in this thesis, as business intelligence (BI) systems, focused on monitoring performance through events that have occurred (metrics), analysed in business perspectives (dimensions).

Each model, regardless of its graphical view, is instantiated in the form of data, but with definitions of this data represented in structures known as metadata. As such, metadata is typically viewed as data about data. The importance of metadata is now recognized in the field of EA, data governance and BI. However, one of the problems is associated with the lack of interoperability between these systems and their metadata, since a transposition between models is necessary, from the definition of the organization to the support to the functioning of the organization at the level of its information system. The lack of interoperability leads to misalignment between the concepts used in EA and their implementation in systems and information, with an impact on the level of the effort for the implementation and maintenance of these systems. This misalignment has a greater impact in the case of BI systems that use business concepts in the form of dimensions and metrics that should be aligned with those used in defining the organization and its strategic objectives, drawn in the EA.

This thesis focuses its investigation on this problem of interoperability in the field of knowledge representation, considering a solution architecture in which metadata is seen as an organizational ontology, modelled in EA, enriched with other concepts (e.g. glossaries, definition of databases), reused in BI systems and allowing its use in user interfaces based on natural language processing. In this way, the effort and complexity of managing EA and BI is reduced, thus making possible a better alignment between the concepts used in performance measurement and the concepts defined in the modelling of the organization, as a contribution to the alignment between the business and information systems. This solution hypothesis thus defines an architecture and approach to support organizational intelligence.

Keywords: Enterprise architecture, business models, business processes, business intelligence, metadata, data lineage, data discovery, knowledge discovery, knowledge management, data governance, ontologies, ontology learning, reasoning.



III. Tabela de conteúdos

1	INTRODUÇÃO	1
1.1	ENQUADRAMENTO	1
1.2	OBJETIVO DE INVESTIGAÇÃO	7
1.3	INVESTIGAÇÃO REALIZADA	8
1.4	CONTRIBUTOS DA INVESTIGAÇÃO	9
1.5	MOTIVAÇÃO	10
1.6	ESTRUTURA DA TESE	10
2	METODOLOGIA DE INVESTIGAÇÃO	11
2.1	ENQUADRAMENTO	11
2.2	ABORDAGEM DE INVESTIGAÇÃO	11
3	ESTADO DA ARTE	13
3.1	INTRODUÇÃO	13
3.2	MODELAÇÃO DA ORGANIZAÇÃO	16
3.2.1	ORGANIZAÇÃO	16
3.2.2	ESTRATÉGIA E MODELO DA ORGANIZAÇÃO	17
3.2.3	OPERACIONALIZAÇÃO ORGANIZACIONAL	20
3.2.4	DESEMPENHO ORGANIZACIONAL	21
3.2.5	MEMÓRIA ORGANIZACIONAL	23
3.3	GESTÃO DE CONHECIMENTO	24
3.3.1	DADOS, INFORMAÇÃO E CONHECIMENTO	24
3.3.2	GESTÃO DE DADOS E CONHECIMENTO	26
3.3.3	REPRESENTAÇÃO DE CONHECIMENTO	29
3.3.3.1	Arquitetura empresarial	29
3.3.3.2	Metadados	37
3.3.3.3	Ontologias	43
3.3.3.4	Web semântica	51
3.3.3.5	Bases de dados	53
3.4	SISTEMAS DE INFORMAÇÃO	58
3.4.1	SISTEMAS DE INFORMAÇÃO	58
3.4.2	ARQUITETURA DE SISTEMAS DE INFORMAÇÃO	61
3.4.3	SISTEMAS INFORMACIONAIS	62
3.4.3.1	Definição	62
3.4.3.2	Arquitetura de componentes <i>business intelligence</i>	64
3.4.3.3	Arquitetura de componentes de dados em <i>business intelligence</i>	67
3.4.3.4	Metodologias <i>business intelligence</i>	70
3.4.4	EXPLORAÇÃO DE CONHECIMENTO	71
3.4.4.1	Definição	71
3.4.4.2	<i>Ontology learning</i>	73
3.4.4.3	<i>Data lineage</i>	74
3.4.4.4	Processamento língua natural	76
3.4.4.5	<i>Reasoning</i>	79
3.4.4.6	<i>Recommender system</i>	79
3.4.4.7	<i>Data visualization</i>	79

3.5	INTEROPERABILIDADE DE SISTEMAS	82
4	CARACTERIZAÇÃO DO PROBLEMA	85
4.1	INTRODUÇÃO	85
4.2	MODELAÇÃO DE ARQUITETURA EMPRESARIAL	87
4.3	IMPLEMENTAÇÃO DE SISTEMA BUSINESS INTELLIGENCE	89
4.4	SISTEMATIZAÇÃO DO PROBLEMA	90
4.5	CAUSA E IMPACTO	91
5	QUESTIONÁRIO A EMPRESAS	93
5.1	INTRODUÇÃO	93
5.2	UTILIZAÇÃO DE ARQUITETURA EMPRESARIAL	93
5.3	UTILIZAÇÃO DE BUSINESS INTELLIGENCE	97
5.4	PERCEÇÃO DO PROBLEMA E ADEQUAÇÃO DA HIPÓTESE DE SOLUÇÃO	99
6	ARQUITETURA PROPOSTA	101
6.1	INTRODUÇÃO	101
6.2	REQUISITOS DERIVADOS DO PROBLEMA	102
6.3	ARQUITETURA DE SOLUÇÃO	104
6.3.1	ARQUITETURA DE REFERÊNCIA	104
6.3.2	<i>METADADOS COMO ONTOLOGIA</i>	105
6.3.2.1	Metadados técnicos	107
6.3.2.2	Metadados de negócio	109
6.3.3	SOLUÇÃO DE <i>BUSINESS INTELLIGENCE</i> ORIENTADA A ARQUITETURA EMPRESARIAL	112
6.3.4	SOLUÇÃO DE ANÁLISE DE METADADOS COMO ONTOLOGIA	117
6.3.4.1	Arquitetura tecnológica	117
6.3.4.2	Arquitetura aplicacional	120
6.3.4.3	Integração de ontologia	126
6.3.4.4	Análise de expressões de negócio	128
6.4	RESULTADOS ESPERADOS	129
7	CASOS DE ESTUDO	131
7.1	INTRODUÇÃO	131
7.2	EMPRESA GESTORA DE INFRAESTRUTURAS DE ÁGUA E ENERGIA	132
7.2.1	ARQUITETURA EMPRESARIAL	132
7.2.1.1	Arquitetura organizacional	132
7.2.1.2	Arquitetura negócio	135
7.2.2	<i>BUSINESS INTELLIGENCE</i>	137
7.2.2.1	Base de dados	137
7.2.2.2	Integração de dados	139
7.2.2.3	<i>Dashboard</i> para exploração de dados	140
7.2.3	ANALISADOR DE EXPRESSÕES DE NEGÓCIO	143
7.3	EMPRESA FINANCEIRA DE CORRETAGEM	143
7.3.1	ARQUITETURA EMPRESARIAL	144
7.3.1.1	Arquitetura organizacional	144

7.3.1.2	Arquitetura negócio	146
7.3.2	<i>BUSINESS INTELLIGENCE</i>	148
7.3.2.1	Base de dados	148
7.3.2.2	Integração de dados	150
7.3.2.3	<i>Dashboard</i> para exploração de dados	151
7.3.3	ANALISADOR DE EXPRESSÕES DE NEGÓCIO	153
7.4	QUESTIONÁRIO DE CASOS DE ESTUDO	153
7.4.1	UTILIZAÇÃO DE ARQUITETURA EMPRESARIAL	153
7.4.2	UTILIZAÇÃO DE <i>BUSINESS INTELLIGENCE</i>	156
7.4.3	EXPERIMENTAÇÃO DA HIPÓTESE DE SOLUÇÃO	158
8	DISCUSSÃO	163
8.1	INTRODUÇÃO	163
8.2	EXPERIMENTAÇÃO DA HIPÓTESE	165
8.2.1	MODELAÇÃO DE SISTEMAS DE <i>BUSINESS INTELLIGENCE</i> A PARTIR DA ARQUITETURA EMPRESARIAL	165
8.2.2	UTILIZAÇÃO DE ONTOLOGIAS EM <i>BUSINESS INTELLIGENCE</i> PARA <i>DESCOBERTA DE CONHECIMENTO</i>	165
8.2.3	UTILIZAÇÃO DE ONTOLOGIAS EM <i>BUSINESS INTELLIGENCE</i> VIA PROCESSAMENTO DE LÍNGUA NATURAL	166
8.3	RESPOSTA A QUESTÕES DE INVESTIGAÇÃO	167
8.4	TRABALHOS RELACIONADOS	169
8.4.1	ONTOLOGIAS EM MODELAÇÃO	169
8.4.2	ONTOLOGIAS EM BASES DE DADOS	170
8.4.3	ONTOLOGIAS EM <i>BUSINESS INTELLIGENCE</i>	171
8.4.4	ONTOLOGIAS EM METADADOS	173
9	CONCLUSÃO	175
9.1	INTRODUÇÃO	175
9.2	PROBLEMA E HIPÓTESE SOLUÇÃO	175
9.3	RESULTADOS ALCANÇADOS E LIMITAÇÕES	178
9.4	PUBLICAÇÕES EFETUADAS	180
9.5	LINHAS DE INVESTIGAÇÃO FUTURAS	181
	REFERÊNCIAS BIBLIOGRÁFICAS	183
	ANEXO A - QUESTIONÁRIO A EMPRESAS SOBRE AE E BI	197
	ANEXO B – QUESTIONÁRIO CASOS ESTUDO	205
	ANEXO C – ARTEFATOS DO PROTÓTIPO EM JAVA	213



IV. Lista de figuras

Figura 1.1: Relação entre modelo de negócio e sistemas de informação.....	3
Figura 1.2: Detalhe da relação entre modelo de negócio e sistemas de informação.....	5
Figura 3.1: Estrutura de modelo de negócio (Orofino, 2011).....	19
Figura 3.2: Estrutura de modelo de negócio (Osterwalder e Pigneur, 2010).....	19
Figura 3.3: Modelo de estratégia (Guimaraes, 2006).....	20
Figura 3.4: Modelo de memória organizacional, adaptado de Walsh e Ungson (1991).....	23
Figura 3.5: Modelos de criação de conhecimento, adaptado de Nonaka (1994).	25
Figura 3.6: Inter-relação de conceitos nas organizações do conhecimento (Angeloni e Fernandes, 2000)	27
Figura 3.7: IBM Information Governance Maturity Model (IBM, 2007; Davis, 2010).....	28
Figura 3.8: Framework de Zachman (Zachman, 2009)	32
Figura 3.9: Modelo de arquitetura TOGAF e ArchiMate (GROUP, 2012)	33
Figura 3.10: Relação entre níveis de arquitetura ArchiMate (GROUP, 2012)	33
Figura 3.11: Business Layer ArchiMate (GROUP, 2012).....	34
Figura 3.12: Application Layer ArchiMate (GROUP, 2012)	34
Figura 3.13: Technology Layer ArchiMate (GROUP, 2012).....	35
Figura 3.14: Framework de Zachman (Zachman, 2009)	36
Figura 3.15: Modelo de arquitetura empresarial	37
Figura 3.16: Modelo IMM vs outros modelos (IMM, 2008)	41
Figura 3.17: Âmbito da especificação IMM (IMM, 2008)	42
Figura 3.18: Tipologia de ontologias segundo John Sowa (Bieman, 2005)	45
Figura 3.19: Pilha tecnológica web semântica (Berners-Lee et al., 2006).....	51
Figura 3.20: Estrutura de uma ontologia OWL (OWL, 2013).....	52
Figura 3.21: Modelo Relacional, adaptado de Astrova (2009)	55
Figura 3.22: Modelo de uma ontologia, adaptado de Astrova (2009)	57
Figura 3.23: Fluxo de processamento informação, adaptado de Laudon e Laudon (2012)..	59
Figura 3.24: Tipologia de sistemas de informação segundo Laudon e Laudon (2000)	59
Figura 3.25: Relação entre Organização e Sistemas de Informação, adaptado de Laudon e Laudon (2004)	62
Figura 3.26: Componentes <i>data warehouse</i> , adaptado de Laudon e Laudon (2012)	64
Figura 3.27: Componentes <i>data warehouse</i> (Kimbali e Ross, 2013).....	64
Figura 3.28: Data Warehouse Data Architecture (Marco, 2013).....	65
Figura 3.29: BI Architecture (Chaudhuri et al., 2011).....	65

Figura 3.30: Gartner business analytic framework (Gartner, 2011).....	66
Figura 3.31: Arquitetura consolidada de <i>business intelligence</i>	66
Figura 3.32: Componentes arquiteturais de uma <i>data warehouse</i> (Berkeley, 2006)	68
Figura 3.33: Modelo multidimensional <i>star schema</i> (Kimbal e Ross, 2013)	70
Figura 3.34: Modelo multidimensional em snowflake (Kimbal e Ross, 2013)	70
Figura 3.35: Processo knowledge discovery (Fayad et al., 1996)	71
Figura 3.36: Processo de ontology learning (Maedche e Staab, 2001)	74
Figura 3.37: Componentes <i>ontology learning</i> (Maedche e Staab, 2001).....	74
Figura 3.38: Gráfico de transformação (Ikeda e Widom, 2009)	75
Figura 3.39: Taxonomia características <i>data lineage</i> (Simmhan et al., 2005).....	75
Figura 3.40: Processador NLP de Nikos Drakos citado por Mote (2002).....	77
Figura 3.41: Estrutura de um <i>dashboard</i> (Elias, 2012)	80
Figura 3.42: Pessoas, domínios de interesse e pontos de vista em <i>data visualization</i> (Buckl et al., 2010)	81
Figura 4.1: Adaptação do modelo de negócio de Osterwalder e Pigneur (2010)	87
Figura 4.2: Modelação de arquitetura empresarial em Archi	88
Figura 4.3: <i>Dashboard</i> informação de gestão em Microsoft PowerBI Desktop	90
Figura 5.1: A quantos anos existe a arquitetura empresarial?.....	94
Figura 5.2: Que unidades orgânicas estão envolvidas na gestão da arquitetura empresarial?	94
Figura 5.3: Que metamodelo é utilizado para arquitetura empresarial?.....	94
Figura 5.4: Que níveis de arquitetura de AE existem no seu modelo?	95
Figura 5.5: Qual a utilização da arquitetura empresarial na organização?	95
Figura 5.6: Que ferramenta é utilizada para AE?.....	95
Figura 5.7: A ferramenta tem integração com outras fontes de artefatos?	96
Figura 5.8: Qual o nível de dificuldade na criação da arquitetura empresarial?.....	96
Figura 5.9: Qual o nível de dificuldade na atualização da arquitetura empresarial?.....	97
Figura 5.10: Qual o principal motivo na dificuldade de gestão da arquitetura empresarial?	97
Figura 5.11: A quantos anos existe a arquitetura de <i>business intelligence</i> ?.....	97
Figura 5.12: Que unidades orgânicas gerem a arquitetura de <i>business intelligence</i> ?	97
Figura 5.13: Que ferramentas são utilizadas para BI?	98
Figura 5.14: Que metamodelo é utilizado para BI?	98
Figura 5.15: Que níveis de arquitetura de BI existem?.....	98
Figura 5.16: Que tipo de metadados existem?.....	98

Figura 5.17: Qual o nível de dificuldade na gestão de <i>business intelligence</i> ?.....	99
Figura 5.18: Qual o principal motivo de dificuldade de gestão de <i>business intelligence</i> ?	99
Figura 5.19: Qual o nível de importância do modelo E como ontologia corporativa?.....	99
Figura 5.20: Qual o nível de importância de classificar e reutilizar dimensões em BI a partir de conceitos e relações em AE?	100
Figura 5.21: Qual o nível de importância de classificar e reutilizar métricas em BI a partir de KPI e objetivos em AE?.....	100
Figura 5.22: Qual o nível de dificuldade na gestão de <i>business intelligence</i> ?.....	100
Figura 6.1: Arquitetura de referência de hipótese de solução.....	105
Figura 6.2: Modelo hipótese de metadados como ontologia	106
Figura 6.3: Metadados técnicos.....	107
Figura 6.4: Metadados de negócio	109
Figura 6.5: Ontologia para arquitetura organizacional.....	111
Figura 6.6: Ontologia para arquitetura de negócio	111
Figura 6.7: Modelo de relação entre arquitetura negócio/organizacional e arquitetura de informação.....	112
Figura 6.8: Abordagem de implementação de <i>business intelligence</i>	112
Figura 6.9: Agregação de Jobs para integração de dados em Microsoft <i>Integration Services</i>	113
Figura 6.10: Processo de integração de dados de uma tabela em Microsoft <i>Integration Services</i>	114
Figura 6.11: Base de dados em Microsoft SQLServer.....	115
Figura 6.12: Modelação de dados em Microsoft PowerBI (Tabelas).....	115
Figura 6.13: Modelação de dados em Microsoft PowerBI (Relações entre tabelas)	116
Figura 6.14: Estrutura de <i>dashboard business intelligence</i>	117
Figura 6.15: Arquitetura técnica de solução em hipótese.....	118
Figura 6.16: Arquitetura aplicacional metadados como ontologia	121
Figura 6.17:Gravar e ler ontologia	123
Figura 6.18: Ler e guardar configurações	123
Figura 6.19: Interpretar e processar ontologia.....	124
Figura 6.20: Análise de expressões de informação de gestão.....	125
Figura 6.21: Gravar e ler expressões de informação de gestão	125
Figura 6.22: Query DML para extrair metadados SQLServer (Tabelas e Campos)	127
Figura 6.23: Query DML para extrair metadados SQLServer (Chaves Estrangeiras).....	127
Figura 6.24: Exemplo de Glossário.....	128
Figura 6.25: Análise de expressões de negócio	128

Figura 7.1: Modelo de ontologia da arquitetura organizacional de empresa gestora de infraestruturas	133
Figura 7.2: Modelo instanciado de ontologia da arquitetura organizacional de empresa gestora de infraestruturas	134
Figura 7.3: Modelo de relação entre arquitetura organizacional e arquitetura de informação de empresa gestora de infraestruturas.....	134
Figura 7.4: Modelo de ontologia da arquitetura negócio de empresa gestora de infraestruturas	135
Figura 7.5: Modelo instanciado de ontologia da arquitetura negócio de empresa gestora de infraestruturas	136
Figura 7.6: Modelo de relação entre arquitetura de negócio e arquitetura de informação de empresa gestora de infraestruturas.....	136
Figura 7.7: Modelo de dados empresa gestora de infraestruturas em PowerBI Desktop ..	137
Figura 7.8: Modelo de relação entre tabelas da empresa gestora de infraestruturas	138
Figura 7.9: Processo ETL empresa gestora de infraestruturas	139
Figura 7.10: Processo ETL empresa gestora de infraestruturas. Exemplo de transformação por tabela.....	139
Figura 7.11: Sheet “Estrutura Organizacional” do <i>dashboard</i> de empresa gestora de infraestruturas	140
Figura 7.12: Sheet “Negócio” do <i>dashboard</i> de empresa gestora de infraestruturas	141
Figura 7.13: Sheet “Proveitos” do <i>dashboard</i> de empresa gestora de infraestruturas	141
Figura 7.14: Sheet “Custos” do <i>dashboard</i> de empresa gestora de infraestruturas.....	142
Figura 7.15: Analisador de expressões de negócio de empresa gestora de infraestruturas	143
Figura 7.16: Modelo de ontologia da arquitetura organizacional de empresa financeira..	145
Figura 7.17: Modelo instanciado de ontologia da arquitetura organizacional de empresa financeira	145
Figura 7.18: Modelo de relação entre arquitetura organizacional e arquitetura de informação de empresa gestora de infraestruturas.....	146
Figura 7.19: Modelo de ontologia da arquitetura negócio de empresa financeira	146
Figura 7.20: Modelo instanciado de ontologia da arquitetura negócio de empresa financeira	147
Figura 7.21: Modelo de relação entre arquitetura de negócio e arquitetura de informação de empresa gestora de infraestruturas.....	148
Figura 7.22: Modelo de dados empresa financeira em PowerBI Desktop	149
Figura 7.23: Modelo de relação entre tabelas da empresa financeira.....	149
Figura 7.24: Processo ETL empresa financeira	150
Figura 7.25: Processo ETL empresa financeira. Exemplo de transformação por tabela	151

Figura 7.26: <i>Sheet</i> “Estrutura Organizacional” do <i>dashboard</i> de empresa financeira.....	151
Figura 7.27: <i>Sheet</i> “Estrutura Negócio” do <i>dashboard</i> de empresa financeira	152
Figura 7.28: <i>Sheet</i> “Proveitos” do <i>dashboard</i> de empresa financeira	152
Figura 7.29: Analisador de expressões de negócio de empresa financeira	153
Figura 7.30: Porque razão não existe arquitetura empresarial (AE)?	154
Figura 7.31: Que unidades orgânicas deviam estar envolvidas na gestão da AE?	154
Figura 7.32: Que níveis de arquitetura deveriam existir no seu modelo em AE?	154
Figura 7.33: Que ferramenta é utilizada para documentar arquiteturas da empresa?	155
Figura 7.34: A ferramenta tem integração com outras fontes?	155
Figura 7.35: Qual o nível de dificuldade na atualização da documentação?	155
Figura 7.36: Qual o principal motivo na dificuldade de documentação?.....	156
Figura 7.37: Que utilização faz da documentação de arquiteturas?.....	156
Figura 7.38: Que níveis de arquitetura de BI existem?.....	156
Figura 7.39: Que unidades orgânicas gerem a arquitetura de BI?	157
Figura 7.40: Que metamodelo é utilizado para BI?	157
Figura 7.41: Que tipo de metadata existe?	157
Figura 7.42: Que ferramentas são utilizadas para BI?	157
Figura 7.43: Qual o nível de dificuldade na gestão de BI?.....	158
Figura 7.44: Qual o principal motivo de dificuldade de gestão de BI?	158
Figura 7.45: Qual a sua opinião sobre as causas e impacto do problema indicado?	159
Figura 7.46: Qual a sua opinião sobre utilização da arquitetura empresarial para definir métricas e dimensões em BI?	159
Figura 7.47: Qual a sua opinião sobre o resultado do questionário ao mercado?	159
Figura 7.48: A arquitetura organizacional apresentada reflete a realidade da sua organização?	159
Figura 7.49: A arquitetura de negócio apresentada reflete a realidade da sua organização?	160
Figura 7.50: A base de dados apresentada reflete as dimensões e factos críticos da sua organização?	160
Figura 7.51: Os <i>dashboards</i> apresentados refletem a estrutura de análises críticas?.....	161
Figura 7.52: O analisador de expressões apresentado corresponde a uma linguagem de interrogação de dimensões e métricas em BI?	161



V. Lista de tabelas

Tabela 3.1: Mapa estratégico para Banca Comercial (Russo, 2006)	22
Tabela 3.2: Relação entre dados, bases dados, modelos de dados e metadados (IMM, 2008)	41
Tabela 3.3: Modelo de relação entre base de dados relacional e ontologia, adaptado de Villanueva-Rosales (2011).....	58
Tabela 3.4: Sistemas operacionais vs sistemas informacionais, adaptado de Mawilmada (2011).....	61
Tabela 6.1: Cenários alternativos de hipótese de solução	102
Tabela 6.2: Conceitos de metadados técnicos	108
Tabela 6.3: Conceitos de metadados de negócio	110
Tabela 6.4: Tecnologias utilizadas em metadados como ontologia	119
Tabela 6.5: Classes Java de acesso a dados	122
Tabela 6.6: Classes Java de exploração de dados.....	122
Tabela 6.7: Análise gramatical de expressão.....	129
Tabela 9.1: Publicações efetuadas durante a investigação.....	180



VI. Lista de siglas

Termo	Descrição
AE	Arquitetura empresarial
AGATE	Atelier de Gestion de l'Arquicheture des Systèmes d'Information et de Communication
ANSI	American National Standards Institute
ASCII	American Standard Code for Information Interchange
ATM	Automatic Teller Machine
BAM	Business Activity Monitoring
BEM	Business Excellence Model
BI	Business intelligence
BIDM	Basic Interoperability Data Model
BPM	Business Process Management
BPMN	Business Process Modelling Notation
BPOKM	Business Process Oriented Knowledge Management
BSC	Balance Scorecard
CAF-C4ISR	Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance) Architecture Framework
CDC	Change Data Capture
CIM	Computation Independent Model
CORBA	Common Object Request Broker Architecture
CPM	Corporate Performance Management
CSV	Comma Separate Value
CWM	Common WareHouse Metamodel
DAM-OIL	DARPA Agent Markup Language – Ontology Interchange
DCL	Data Control Language
DCMI	Dublin Core Metadados Initiative
DDL	Data Definition Language
DEA	Data Envelopment Analysis

DM	Data Mart
DML	Data Manipulation Language
DODAF	Department of Defense Architecture Framework
DSS	Decision Support System. Equivalente a SAD.
DTD	Document Type Definition
DTL	Data Transaction Language
DW	Data Warehouse
E2AF	Extended Enterprise Architecture Framework of Institute for Enterprise Architecture Development
EAI	Enterprise Architecture Integration
EBCDIC	Extended Binary Coded Decimal Interchange Code
E-GIF	eGovernment Interoperability Framework
EIS	Executive Information System. Equivalente a SIE.
ERP	Enterprise Resource Planning
ESM	Enterprise System Metamodel
ETL	Extract, Transform and Load
FEAF	Federal Enterprise Architecture Framework
FIPA	Foundation for Intelligent Physical Agents
IAF	Integrated Architecture Framework da CapGemini
IEEE	Institute of Electrical and Electronics Engineers
IIF	Australia Government Information Interoperability Framework
IMM	Information Metadados Metamodel
ISO	International Standard Organization
ITPMM	IT Portfolio Management Metamodel
KPI	Key Performance Indicator
LEAN	LEAN Manufacturing Quality Tools
MDA	Model Driven Architecture
MDC	Meta Data Coalition
MDIS	Metadados Interchange Specification

MDX	Multidimensional Expression
MEAF	Metis Enterprise Architecture Framework
MIS	Management Information System. Equivalente a SIG.
MME	Managed Metadados Environment
MODAF	Ministry of Defense Architecture Framework
MOF	Meta Object Facility
NAF	NATO Architecture Framework
ODBC	Open Database Connectivity
OIL	Ontology Interchange Language
OIM	Open Information Model
OLAP	On-Line Analytical Processing
OLTP	OnLine Transaction Processing
OMG	Object Management Group
OML	Ontology Markup Language
OWL	Web Ontology Language
PIM	Plataform Independent Model
PMM	Performance Measurement Matrix
PP	Performance Prism
PSM	Plataform Specific Model
R&DM	Results and Determinants Matrix
RDF	Resource Description Language
RDF-S	Resource Description Language Schema
SAD	Sistema de Apoio à Decisão. Equivalente a DSS.
SFA	Sales Force Automation
SGBD	Sistema de Gestão de Base de Dados
SHOE	Simple HTML Ontology Extension
SI	Sistemas de Informação
SIE	Sistema de informação executiva. Equivalente a EIS

SIG	Sistema de informação de gestão. Equivalente a MIS
SKOS	Simple Knowledge Organization System
SMART	Strategic Measurement and Reporting Techniques
SPARQL	Protocol and RDF Query Language
SQL	Structured Query Language
SWAD	Semantic Web Advanced Development
SWOT	Strength, Weaknesses, Opportunities and Threats
SWRL	Semantic Web Rule Language
TEAF	Treasury Enterprise Architecture Framework
TI	Tecnologia de Informação
TOGAF	The Open Group Architecture Framework
UML	Unified Modelling Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
WBS	Work Breakdown Structure
XMI	XML Metadados Interchange
XML	Extensible Markup Language
XMLS	XML Schema
XOL	XML-Based Ontology Exchange Language

1 Introdução

Neste capítulo apresenta-se o enquadramento da tese (sub-capítulo 1.1) como um resumo sobre a necessidade de se implementar modelos de inteligência organizacional assentes em informação sobre a própria organização, mediante a relação estabelecida entre arquitetura empresarial e *business intelligence*. Com base neste enquadramento apresenta-se o objetivo da investigação (sub-capítulo 1.2), a investigação realizada (sub-capítulo 1.3) e os contributos da investigação para o domínio de conhecimento (sub-capítulo 1.4). Por fim, apresenta-se a motivação para a realização desta investigação (sub-capítulo 1.5) e a estrutura da tese (sub-capítulo 1.6).

1.1 Enquadramento

Rodrigues (2014) posiciona a importância da arquitetura empresarial (AE) nas organizações pelo facto de permitir um alinhamento entre sistemas de informação e o negócio, além de permitir reduzir a complexidade da gestão interligada de várias componentes da organização. No entanto, o mesmo autor conclui que *“a análise do valor das Arquiteturas Empresariais é indiscutivelmente cada vez mais importante para as organizações face à dificuldade destas em justificar o investimento significativo que é necessário para conceber, implementar e manter um projeto desta natureza”* (Rodrigues, 2014).

Na nossa investigação consideramos a AE como referencial base para a investigação, pelo seu potencial de permitir um maior alinhamento entre sistemas de informação e negócio, tal como referido por Rodrigues (2014), mas focado no problema de reutilização de conceitos entre a definição do modelo da organização capturado na AE e a monitorização do seu desempenho através de informação disponível no sistema de *business intelligence* (BI). Como tal, acrescentamos desta forma uma justificação para a utilização da AE enquanto valor para as organizações, pois a integração entre os dois sistemas, garante que a AE possa ser vista como um ativo de informação que permite capturar a representação de conhecimento da própria organização reutilizando depois essa informação. Como tal, analisamos diversos modelos de gestão no sub-capítulo 3.2 (modelação das organizações) para destacar as várias componentes que depois apresentamos como parte da análise do domínio da AE no sub-capítulo 3.3.3.1 (arquitetura empresarial) que culmina com um referencial proposto na investigação na Figura 3.15 no final desse sub-capítulo. Reconhecemos desta forma que existem várias formas de modelar a organização, mas como ponto de partida, consideramos que AE permite integrar os vários modelos da organização numa visão holística. Consideramos igualmente que a AE corresponde a uma forma de representar conhecimento sobre a própria organização, sendo que segundo Sowa (2000a, 2000b), Gruber (1993) e Noy e McGuinness (2001), representar conhecimento corresponde a criar uma ontologia, sendo como tal a AE uma ontologia. Esta visão é reforçada por The Open Group [GROUP] (2012, 2016) que descrevem a AE como a descrição formal de um sistema que permite raciocinar sobre a sua estrutura e comportamento, o que está alinhado com a definição da organização como um sistema e a sua captura como representação de conhecimento como sendo uma ontologia. No nosso entendimento, a própria definição da AE funciona ela própria como uma forma de metadados de negócio para a própria organização, pois permite detalhar que conceitos existem, como se relacionam e onde estão localizados em termos de sistema de gestão de informação.

Considerando assim que a AE, vista como ontologia, permite definir a organização através de um conjunto de conceitos (estrutura) e suas relações (semântica) capturadas em modelos, a nossa investigação foca a necessidade de reutilizar essa informação por parte do sistema de *business intelligence (BI)*, também designado por sistema informacional, para que a monitorização de desempenho utilize a mesma linguagem utilizada no modelo da organização representado na arquitetura empresarial. Como tal, investigamos no domínio da arquitetura de sistemas de informação no sub-capítulo 3.4 (sistemas de informação) o posicionamento do sistema de *BI*, separando este tipo de sistemas informacionais do que é considerado como sistemas operacionais:

- Arnet et al. (2000) e Moss e Altre (2003) concordam que é um tipo de sistema que permite combinar dados, informação e conhecimento acerca do ambiente interno e externo do negócio com o objetivo de suportar a tomada de decisão, facilitando assim a gestão da dinâmica da organização no seu meio ambiente e das suas vantagens competitivas. Nesta definição torna-se claro que o objetivo é utilizar dados que representam o conhecimento da organização e do meio onde se insere. No entanto, consideramos na nossa investigação, que a tomada de decisão sobre a estrutura e semântica da organização é efetuada na AE para depois se monitorizar o desempenho, através dos mesmos conceitos em BI;
- Rossini e Palmisano (2003) referem que estes sistemas combinam conceitos de *data warehouse*, *data mart* e OLAP, com sistemas específicos de EIS, *data mining* e DSS. Neste caso o autor enuncia componentes que fazem parte de um sistema de *BI*, tornando clara a abrangência do conceito em termos de tipo de sistemas que dele faz parte;
- Para Gangadharan e Sundaravalli (2004) o sistema de BI corresponde a uma arquitetura que tem como objetivo a integração de vários componentes ao nível de bases de dados e aplicações operacionais, consolidadas para dar acesso a dados organizados com sentido adequado para vários níveis de tomada de decisão consistentes, no contexto da monitorização de desempenho da estratégia e operações. Para estes autores, BI representa uma nova disciplina de gestão em que os dados são tratados como recursos empresariais, alertando para o processo de monitorização da organização, que remete igualmente para uma necessidade de interligação entre a definição da organização ao nível de AE e sua monitorização ao nível de BI, com base nos mesmos conceitos.

Pelo facto da AE poder estar suportada em linguagens como ArchiMate e em ferramentas específicas, e a representação de informação em base de dados de BI utilizar a linguagem DDL (*Data definition language*) e utilizar ferramentas específicas, permite-nos constatar que ambas têm em comum estruturas de metadados, razão pela qual consideramos este conceito como aspeto crítico da nossa investigação enquanto parte da hipótese de solução. De acordo com NISO (2004), os metadados correspondem a informação estruturada que descreve, explica, localiza e permite gerir os recursos de informação. Marco (2013) acrescenta que essa informação é guardada em aplicações informáticas e que corresponde a conhecimento sobre a organização, podendo separar-se em metadados técnicos e de negócio. Por outro lado, para Imnon et al. (2010), no domínio de BI, os metadados são a sua componente central e crítica. Como tal, na nossa investigação consideramos que os metadados permitem obter os conceitos descritos em AE, base de dados e glossários, com níveis de detalhe diferentes em cada sistema, sendo que o foco da nossa investigação foi centrado em encontrar uma solução para integrar os vários metadados e criar uma ontologia

como resultado da integração entre os vários conceitos. Por este motivo recorreremos na investigação ao domínio de *ontology learning* (Azman et al., 2011; Maedche e Staab, 2001). Esta visão de alinhamento entre metadados e ontologias é igualmente referido por Scilia (2006) ao indicar a tendência multidisciplinar de combinar metadados com ontologias para partilha de semântica ao nível de metadados como parte da interoperabilidade entre sistemas. Por outro lado, Mochol (2009) refere que a interoperabilidade é um tema abordado em várias investigações em Web semântica, motivo pelo qual enquadrámos o tratamento das ontologias dentro do domínio da Web semântica.

Face ao contexto acima descrito, consideramos na nossa investigação a visão de uma organização como um sistema que é representado sob a forma de conhecimento sobre si própria como se fosse uma ontologia residente em metadados e com o potencial de ser reutilizada para alinhamento entre a definição do modelo de negócio ao nível de AE e a monitorização do desempenho organizacional através de BI. Como tal, como apresentado na Figura 1.1, consideramos que uma organização (empresa ou outra forma de entidade) é modelada em sistemas de representação de conhecimento (AE e bases de dados), sendo que a semântica e regras de negócio são implementadas em sistemas de processamento de informação (operacionais e informacionais). A informação armazenada em sistemas de representação de conhecimento é utilizada por sistemas de exploração de conhecimento (e.g. *dashboard, knowledge & data discovery*). A conjugação entre sistemas informacionais e sistemas de exploração de conhecimento é designado nesta tese por sistema de BI.

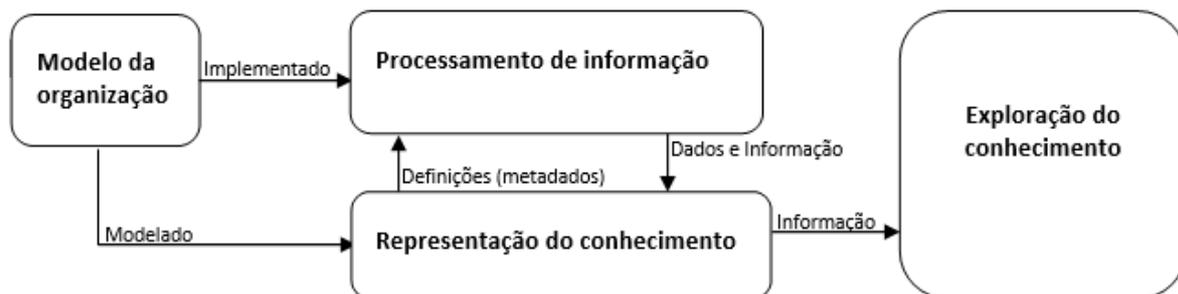


Figura 1.1: Relação entre modelo de negócio e sistemas de informação

Uma organização pode assim ser vista como um sistema, e como tal, passível de ser representada através de um modelo que permita capturar a sua estrutura e semântica. De acordo com Caldeira (2009), a representação de conhecimento pode estar associada a conceitos, enquanto representação de um objeto ou acontecimento que entra na composição de um sistema, que pode ser representada e transferida para o modelo de dados, de onde a importância da arquitetura de informação. Nas organizações, existem assim dados que são transformados em informação pela utilização de sistemas no contexto de processos de negócio, sendo que o conhecimento resulta da ação dinâmica das pessoas na utilização desta informação. Os dados são capturados, registados e tratados sob vários formatos e em vários sistemas, sendo necessário gerir a estrutura da informação (definição) e a semântica (regras de negócio e significado). Esta estrutura e semântica depende de uma classificação inicial (taxonomia) complementada com uma classificação do utilizador (anotações).

Para melhor sistematizarmos o entendimento dos dados, consideramos que existem duas naturezas distintas:

- **Dados capturados e processados pelos sistemas:** derivam da atividade da empresa de acordo com o seu negócio e estratégia, que permitem capturar e tratar informação sobre vários temas (e.g. clientes, fornecedores, produtos, serviços, recursos humanos, contratos, transações). Estes dados têm estruturas e semânticas para suporte a processos de negócio onde se conjugam recursos humanos, recursos materiais e sistemas de informação em torno das atividades da empresa, sendo estruturados em bases de dados. São modelados em fases de desenvolvimento de sistemas nas etapas de análise e desenho das soluções, focando-se normalmente na necessidade de definir as estruturas de tabelas em bases de dados e ficheiros de configuração a serem utilizados por programas que compõem os sistemas, designados por aplicações informáticas;
- **Metadados organizacionais:** estão associados à definição da própria empresa enquanto sistema. Cada conceito (e.g. clientes, fornecedores, processos, estruturas orgânicas, pessoas, mercado, tecnologia) e relações entre si, correspondem igualmente a estruturas de dados que são representadas e armazenadas. Por exemplo, para se ter uma estrutura de dados de clientes, existe antes uma estrutura de dados que define o que é um cliente bem como as regras de relação com produtos e canais comerciais. Estes tipos de dados são utilizados para definição das bases de dados, mas também para reflexão estratégica, planeamento e aferição sobre alinhamento entre a estratégia e sua operacionalização, sendo nesse caso utilizados para definição de modelos da empresa, utilizando-se ou não AE para o efeito. Estes dados residem em repositórios estruturados, não estruturados e documentais, sendo processados por ferramentas de modelação de sistemas, modelação de bases de dados, modelação de AE e modelação de metadados de BI, além de vários documentos descritivos ou de especificação de sistemas, incluindo glossários. São criados em fase de planeamento estratégico, modelação de processos ou modelação de sistemas de informação, sendo reutilizados em sistemas de BI.

Apesar de serem duas naturezas distintas, existe uma estreita correlação entre as mesmas, visto que os dados capturados e processados nas aplicações informáticas dependem das definições dos metadados organizacionais. Nesta correlação torna-se relevante a importância da classificação de termos e conceitos (taxonomia) para garantir interoperabilidade entre os vários sistemas. Daí a importância dos metadados organizacionais e da sua classificação para permitir reutilização entre vários sistemas. No entanto, pela sua dispersão de representação em vários sistemas, torna-se um problema que obriga a uma solução que permita obter uma visão holística para correlacionar e integrar este conhecimento. Em particular, o sistema de BI, como tem por objetivo suportar a tomada de decisão, depende deste alinhamento com conceitos definidos estrategicamente no sistema de AE. O entendimento desta dependência torna-se mais evidente ao se decompor o modelo da Figura 1.1 nas suas várias componentes apresentadas na Figura 1.2, onde se representa os vários tipos de sistemas envolvidos, sendo claro que o modelo da organização é implementado de forma diferente em vários tipos de sistemas, visto que os sistemas operacionais são normalmente constituídos por várias aplicações informáticas de acordo com os processos de suporte ao negócio. Este modelo permite assim enquadrar os vários domínios e suas relações para contextualizar o foco da investigação.

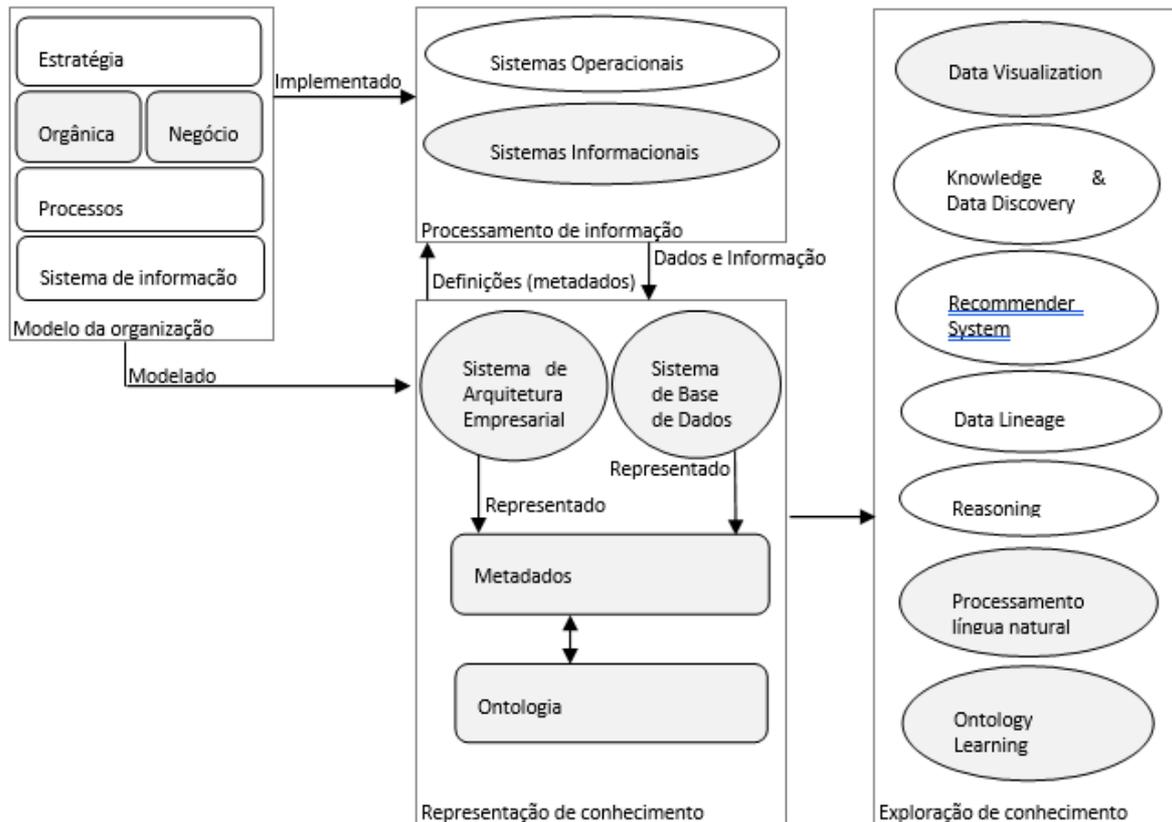


Figura 1.2: Detalhe da relação entre modelo de negócio e sistemas de informação

Esta tese foca-se nesta dependência, vista como um problema de interoperabilidade, explorando o domínio das ontologias como hipótese de solução, mas considerando somente a relação entre o sistema de AE (onde nasce a modelação enquanto identificação e classificação dos conceitos) e o sistema de BI (onde se trata a informação para monitorização de desempenho), sendo como tal o início e fim da informação crítica. Este enfoque é considerado ao nível de unificação de metadados, através de ontologias, com o objetivo de criar uma visão única onde se possa identificar métricas e dimensões a partir de conceitos de negócio relacionados com conceitos de dados e explorar essa ontologia para pesquisa de metadados utilizando processamento de língua natural.

A tese delimita assim o seu estudo ao problema de monitorização do desempenho organizacional através de alinhamento necessário entre o modelo definido da organização e a sua instanciação em sistemas de informação para tomada de decisão. A partir do alinhamento entre conceitos representados na arquitetura empresarial (modelo) e conceitos de dimensões e métricas utilizados no sistema de *business intelligence* (instanciação do modelo), investigamos a relação entre os dois sistemas comparando a representação da organização enquanto modelo de informação, formas de exploração deste modelo de informação e formas de reutilizar este modelo enquanto ontologia que permite igualmente a sua reutilização em sistemas de processamento de língua natural.

Para o efeito, consideramos a hipótese de utilizar a AE vista como ontologia, enriquecida com glossários e modelos de dados. Desta forma, podemos identificar as métricas, dimensões e suas relações, reutilizadas e visualizadas em sistemas de BI e passíveis de serem a base para interpretar a linguagem organizacional neste modelo de ontologia de domínio.

Na nossa investigação consideramos assim um subconjunto de sistemas representado na Figura 1.2:

- **Sistema de arquitetura empresarial (AE):** faz parte dos sistemas de informação enquanto forma de modelação e planeamento da empresa. Permite modelar o negócio em níveis arquiteturais de estratégia, organizacional, negócio, processos e sistemas de informação (aplicações, tecnologia e informação). Utiliza um modelo de representação de conceitos, propriedades e relações, registados como metadados em modelos proprietários;
- **Sistema de bases de dados:** corresponde a repositórios dos dados da empresa para qualquer tipo de sistema, quer seja operacional, informacional ou mesmo ferramentas de modelação. Podem ter formas de representação relacionais, multidimensionais ou semânticas (utilizando linguagens OWL/RDF), utilizando-se linguagens específicas reutilizadas entre várias implementações (sistemas de gestão de bases de dados distintos podem fazer pequenas alterações na sintaxe, mas mantendo o mesmo modelo base da linguagem), como é o caso do SQL (*Structured Query Language*) ou SPARQL (*Protocol and RDF Query Language*). São implementados com conceitos chave de tabelas, campos e relações;
- **Metadados:** correspondem a definições utilizadas pelos próprios sistemas. Podem ser vistos como metadados técnicos (e.g. identificação e relação entre tabelas, campos, programas, cadeias de processamento) e de negócio (descrição e relação entre conceitos de negócio como contrato, cliente e produtos, incluindo fórmulas de cálculo). São utilizados em sistemas de gestão de base de dados para definir estruturas de dados via linguagem DDL (*data definition language*), em soluções de BI ao nível de ETL (*extract, transform and load*) e ao nível de exploração de informação de gestão, em *Packages* aplicacionais (configuração de regras de negócio) ou mesmo em ferramentas de análise de linguagens de programação para extrair regras de negócio em programas. As estruturas de metadados utilizam normalmente formas de representação proprietárias, apesar de existirem iniciativas de reutilização de metamodelos como é o caso do *Information Management Metamodel* [IMM] (2008) e DCMI (*Dublin Core Metadata Initiative*);
- **Ontologias:** com o advento da *web* semântica, criaram-se *normas* para representação de dados por ontologia via especificações OWL (*Ontology Web Language*) e SWRL (*Semantic Web Rule Language*), e com exploração via SPARQL. Permitem representar a informação de forma flexível com classes, propriedades e relações a partir da qual suportam processos de inferência para implementar raciocínio e aprendizagem;
- **Sistemas de *business intelligence* (BI):** corresponde ao conjunto de sistemas de exploração de conhecimento, complementado com os sistemas informacionais ao nível de processamento de informação, por nossa definição. Na tese, focamos os sistemas informacionais, *data visualization*, processamento de língua natural e *ontology learning*.

Estes sistemas estão relacionados com vários domínios de investigação, razão pela qual esta tese aborda os seguintes temas multidisciplinares, detalhados no capítulo 3 (estado da arte):

- **Modelação das organizações:** análise dos conceitos da organização e suas relações, para entender a sua estrutura e a sua semântica, além da lógica de funcionamento ao nível de operacionalização e a forma como efetua a gestão do seu desempenho. Permite identificar a correspondência entre estes conceitos e as dimensões e métricas utilizadas em informação de gestão. Este domínio da gestão é utilizado para se analisar a forma como as ontologias podem ser utilizadas como metamodelos da organização;
- **Gestão de conhecimento:** relação entre dados, informação e conhecimento, separando o que são dados de processamento de negócio e dados que representam a organização. Permite identificar os sistemas onde se representa este conhecimento ao nível de AE, metadados, ontologias, web semântica e bases de dados;
- **Sistemas de informação:** análise da arquitetura de sistemas de informação entre aplicações, tecnologia e informação, para entendimento da forma como os sistemas implementam e suportam o modelo da organização. Separa-se as aplicações entre sistemas operacionais (associados a processos de negócio), sistemas informacionais (que capturam, transformam e consolidam dados em informação para tomada de decisão), sistemas de exploração de conhecimento (*dashboard*, *data discovery*, *data governance*, *ontology learning*, *data lineage*, inferência, *recommender system* e processamento de língua natural) e sistemas de modelação. É considerado na nossa investigação que os sistemas de *BI* correspondem ao conjunto de sistemas informacionais e sistemas de exploração de conhecimento.

1.2 Objetivo de investigação

Considerando que o modelo da organização é definido como uma arquitetura empresarial utilizando-se conceitos e relações entre conceitos, e a monitorização do desempenho é realizada com suporte a sistemas de *business intelligence* suportado em dimensões e métricas, a questão central enquanto objetivo da investigação é a seguinte:

Como utilizar metadados como ontologias para garantir a interoperabilidade entre sistema de arquitetura empresarial e sistema de business intelligence?

O foco da investigação é a interoperabilidade, utilizando a estrutura de metadados do sistema de arquitetura empresarial, enriquecida com o sistema de base de dados de *business intelligence* e glossário de termos da organização, para integrar numa ontologia. Como tal, com base na questão central, estruturaram-se três questões complementares que nos conduzem à questão central de investigação:

Como implementar um sistema de business intelligence a partir do modelo definido numa arquitetura empresarial?

Qual a relação entre conceitos definido numa arquitetura empresarial e dimensões e métricas em business intelligence?

Como reutilizar a ontologia criada para responder a questões de utilizadores via processamento de linguagem natural, no contexto de knowledge discovery?

1.3 Investigação realizada

Face ao objetivo de investigação, foram realizadas as seguintes atividades:

- Revisão da literatura ao nível de modelação das organizações, gestão de conhecimento (arquitetura empresarial, bases de dados, metadados, ontologias) e sistemas de informação (arquitetura de sistemas e *business intelligence*). As conclusões estão detalhadas no capítulo 3, servindo de base para a caracterização do problema no capítulo 4;
- Elaboração e análise de questionário para nove organizações, em sectores de atividade distintos (banca, seguros, telecomunicações e serviços móveis), para identificar a maturidade na utilização de sistemas de arquitetura empresarial e sistemas de *business intelligence*, além de se questionar a perceção sobre a relação entre os dois tipos de sistemas. Para suportar a comunicação com as organizações, foi elaborado um *website* *eaintelligence*¹. As conclusões estão detalhadas no capítulo 5;
- Criação de um modelo de arquitetura empresarial em ArchiMate, com os principais conceitos e relações, de acordo com um modelo proposto nesta tese, para ser testado em casos de estudo. O modelo é apresentado no capítulo 6;
- Elaboração de um modelo de *business intelligence* implementado com a ferramenta Microsoft BI (SQLServer, Integration Services e PowerBI Desktop) com uma estrutura de dados e forma de exploração de métricas e dimensões, a ser testado em casos de estudo. O modelo é apresentado no capítulo 6;
- Desenvolvimento de protótipo em ApacheJENA (OWL, inferência, SPARQL), OpenNLP e Java, para permitir criar uma ontologia a partir de glossários, DDL e ArchiMate, como base para experimentar analisar expressões de *business intelligence* no formato “ANALISAR <lista de métricas> DE <lista de domínios de informação> POR <lista de dimensões> PARA <lista de restrições>”, em idioma Português, a partir da qual se possa analisar a estrutura e semântica de cada palavra na linguagem organizacional criada como ontologia a partir do modelo da organização. O protótipo é apresentado no capítulo 6;
- Teste dos modelos de arquitetura empresarial, *business intelligence* e protótipo de analisador de expressões e ontologia, em duas organizações em sectores de atividade distintos, utilizadas como casos de estudo. No seguimento dessa experimentação, cada organização respondeu a um questionário sobre a aplicabilidade da relação entre arquitetura empresarial e *business intelligence*. O resultado é apresentado no capítulo 7, seguido da discussão final sobre a investigação realizada (sub-capítulo 8.2).

¹ <https://eaintelligence.wordpress.com>

1.4 Contributos da investigação

De acordo com a pesquisa realizada, não existem investigações totalmente equivalentes, tal como apresentado nos trabalhos relacionados (sub-capítulo 8.4). O que existe são aproximações parcelares ao problema, que se podem agrupar da seguinte forma:

- As ferramentas de *business intelligence*, *open source* ou comerciais, permitem identificar as métricas e dimensões a partir da definição dos campos, pelo seu domínio de valores ou nomenclaturas, mas não utilizam ontologias para o efeito;
- As ontologias têm sido utilizadas para permitir explorar bases de dados relacionais a partir de SPARQL, utilizando um modelo de tradução, ou permitindo criar um modelo RDF/OWL a partir de um modelo relacional. Existem inclusive sistemas de gestão de bases de dados baseados em RDF/OWL, *open source* ou comerciais. Mas não estabelecem a relação entre modelo da organização e *business intelligence*;
- O processamento de língua natural tem sido utilizado para permitir criar interfaces com utilizadores para gerar *queries* em MDX/OLAP e SQL. Existem casos de linguagens estruturadas sobre ontologias, mas sem dependerem da definição inicial da ontologia a partir da arquitetura empresarial e sem integração de glossários e modelos de dados.

Considerando o aspeto multidisciplinar da investigação e a ausência de investigações equivalentes, os contributos podem ser apresentados no domínio da aplicabilidade de ontologias nas organizações, detalhado no seguinte:

- Posicionamento de metadados como ontologias para interoperabilidade entre o sistema de *business intelligence* e o sistema de arquitetura empresarial, mantendo a arquitetura empresarial o papel de identificar, classificar e relacionar conceitos chave enquanto sistema central para definição de dos conceitos e relações principais de uma ontologia corporativa;
- Análise da exequibilidade e limitações de utilização de ontologias como forma de representar metadados corporativos, considerando integração por *ontology learning* e utilização de inferência como forma de *data lineage*;
- Definição de abordagem metodológica de implementação de sistemas de *business intelligence* face à identificação e classificação de dimensões e métricas a partir do sistema de arquitetura empresarial para garantir o alinhamento de conceitos e permitindo a exploração da arquitetura empresarial sob a forma de *dashboard* em *business intelligence*;
- Forma de utilização de processamento de língua natural sobre ontologias, para analisar expressões da organização que permitam criar formas de interação com sistemas de *business intelligence* e sistemas de arquitetura empresarial.

Por outro lado, esta investigação traz contributos de revisão de literatura em vários domínios, em particular no papel das ontologias e sua correlação com modelação das organizações e sistemas de informação. Por este motivo a revisão de literatura (capítulo 3) é apresentada em termos de definições e metodologias para cada domínio relacionado com a investigação, reservando-se no entanto os trabalhos relacionados para serem apresentados como parte da discussão (capítulo 8) após se detalhar a investigação realizada.

1.5 Motivação

A motivação para esta investigação tem várias origens:

- Na minha tese de mestrado (Guimaraes, 2006) abordei o problema para tentar encontrar uma solução delimitada à integração entre a arquitetura empresarial e sistemas de *workflow* utilizando como solução o domínio dos agentes inteligentes. Esta investigação é assim uma continuidade para poder detalhar o estudo destes modelos e sua problemática, agora acrescido da visão de ecologia de informação (Caldeira, 2009) enquanto metadados e explorando a utilização de ontologias no domínio específico de *business intelligence* (Saias et al., 2012);
- Pela importância do tema ao nível de tendências tendo por base a necessidade de integração entre arquitetura empresarial e *business intelligence* como parte do conceito de *data governance* e da exploração de dados em *business intelligence* via processamento de língua natural;
- Face ao meu interesse e experiência profissional de desenho e implementação de arquitetura empresarial, sistemas de informação operacionais e sistemas de *business intelligence*. A este nível, sempre senti a necessidade de integrar os dados definidores da organização num tipo de solução que permita acelerar a captura de requisitos para implementação de sistemas de *business intelligence*, que valorize o modelo de arquitetura empresarial, que facilite a gestão estratégica, que permita o suporte a auditorias de informação e possibilite a estruturação de soluções de *data governance*.

1.6 Estrutura da tese

A tese está estruturada da seguinte forma:

- **Capítulo 2:** enquadramento epistemológico para reflexão e conclusão sobre a metodologia de investigação utilizada na tese com a devida justificação sobre a sua adequação ao problema em investigação;
- **Capítulo 3:** identificação e caracterização dos domínios de conhecimento envolvidos no problema e hipótese de solução;
- **Capítulo 4:** descrição do problema considerando a definição dos conceitos relevantes de acordo com o estado da arte, a causa e impacto do problema para as organizações e o papel dos sistemas de informação no problema identificado;
- **Capítulo 5:** conclusões sobre o questionário a nove organizações, relativo à maturidade na utilização de arquitetura empresarial e *business intelligence*;
- **Capítulo 6:** descrição da hipótese e arquitetura de solução para experimentação;
- **Capítulo 7:** apresentação dos resultados de experimentação da hipótese de solução em dois casos de estudo;
- **Capítulo 8:** discussão sobre os resultados de investigação e análise comparativa face a outras referências de solução no mesmo domínio de problema e hipótese de solução;
- **Capítulo 9:** conclusão sobre o problema, adequação da solução, resumo de publicações efetuadas e linhas de investigação futuras no seguimento desta tese.

2 Metodologia de investigação

Neste capítulo apresenta-se o enquadramento teórico sobre metodologias de investigação (sub-capítulo 2.1) para fundamentar a abordagem seguida (sub-capítulo 2.2) considerando que a investigação envolve questionários, fundamentação teórica, protótipo e modelos como proposta de modelo de solução e experimentação em casos de estudo.

2.1 Enquadramento

A investigação é um estudo para obter determinado tipo de informação. Pode ser classificado enquanto estudo exploratório, descritivo ou explanatório (Robson, 1993; Yin, 2013). Pode-se ainda agrupar a prática de investigação de acordo com abordagens ou paradigmas, cada uma com métodos específicos, tal como referido por Cooper et al. (2003) e a seguir apresentado:

- **Abordagem qualitativa:** detalhar uma visão de determinadas situações para compreensão de processos e inter-relações entre objetos de investigação, normalmente em fenómenos sociais e culturais, mas igualmente aplicável em tecnologias de informação. A revisão da literatura poderá ser efetuada à medida que o processo de investigação avança, fazendo parte do processo o próprio investigador que procura respostas às questões de investigação. Desdobra-se em métodos específicos como é o caso de *case study*, *actor network*, *action research* e *ground theory*;
- **Abordagem quantitativa:** explicar relações causais, generalizações e previsão de acontecimentos, normalmente no contexto de fenómenos naturais. A investigação é mais descritiva/causal procurando resultados concretos através de teste de hipóteses com o objetivo de quantificar dados e generalizar resultados a partir da amostra. A revisão da literatura deverá estar completa antes do início do estudo, sendo que o investigador é independente do processo de investigação. Desdobra-se em métodos específicos como é o caso de questionários, trabalhos em laboratório e modelações matemáticas.

No entanto, as abordagens de investigação qualitativas e quantitativas não são mutuamente exclusivas, mas sim complementares como refere Villiers (2005) e Saunders et al. (1987), podendo recorrer-se a uma abordagem qualitativa num trabalho exploratório que sirva para uma posterior investigação quantitativa num estudo descritivo.

2.2 Abordagem de investigação

Face ao enquadramento relativo a metodologias de investigação, seguimos uma abordagem mista visto que utilizamos métodos quantitativos (questionários, trabalho em laboratório e modelação) e qualitativos (questionários, *action research*, *grounded theory* e *actor-network theory*), pelo facto da investigação envolver definição de modelos, avaliação de percepção geral das organizações face ao problema e hipótese de solução, desenvolvimento de protótipo para experimentação, aplicação prática do protótipo em empresas reais e avaliação da adequação da hipótese pela percepção específica das empresas envolvidas nos casos de estudo.

Como tipo de investigação, consideramos um estudo que analisa relações entre dois ou mais conceitos (e.g. arquitetura empresarial, metadados, *business intelligence* e ontologias), sendo como tal um estudo exploratório pelo facto de abordarmos um problema pouco estudado e estudo descritivo pois descrevemos em detalhe cada domínio em investigação.

Utilizamos casos de estudo pois segundo Robson (1993), corresponde a uma abordagem empírica de um fenómeno particular utilizando múltiplas fontes (e.g. documentos, artefactos, observações) permitindo uma visão holística do fenómeno. Para Myers (1997), além de sondagens e experiências de laboratório, o caso de estudo é o método mais utilizado em sistemas de informação. Como tal, para permitir aferir sobre a causa e impacto do problema, além de permitir obter dados para teste da hipótese de solução, foram utilizados casos de estudo em sectores de atividade empresarial distintos para analisar o modelo de organizações, linguagens específicas de negócio e sua relação com conceitos de dimensões e métricas.

No caso dos questionários, estes foram utilizados para permitir obter a perceção da relevância da relação entre arquitetura empresarial e *business intelligence*, além de verificar a pertinência do problema e a adequação da hipótese de interoperabilidade proposta.

A utilização de *action research* deveu-se ao facto de ser uma abordagem que parte da experiência prática e do envolvimento do investigador como refere Stowell et al. (1997). De acordo com Locke (2001), esta abordagem é utilizada em processos de investigação cíclicos e com várias fases em que o desenvolvimento laboratorial da hipótese de solução envolve o investigador de acordo com a sua experiência. Na tese, desenvolveu-se em laboratório um modelo de arquitetura empresarial, um modelo de sistema de *business intelligence* e uma solução programada em linguagem *Java* para integração e gestão da ontologia, para ser depois aplicada em cada caso de estudo, tendo por base a experiência do investigador em colaboração com responsáveis das organizações envolvidas nos casos de estudo.

No caso de *grounded theory*, de acordo com Villiers (2005), é normalmente aplicado em trabalhos de investigação cujo enfoque seja o desenho de modelos teóricos tendo por base a análise das melhores práticas, permitindo desta forma, encontrar uma explicação e compreensão do problema enunciado. Na tese foram desenhados modelos com a estrutura de conceitos e relações chave a considerar num metamodelo de uma organização a partir de ArchiMate e DDL como referenciais, vistos como metadados técnicos e de negócio, aplicados depois em casos de estudo.

A utilização da abordagem *actor-network theory* permite concretizar um método de análise de sistemas de informação, considerando-os como sistemas sociotécnicos, sendo que *actor-network* são pessoas, objetos e organizações (Law, 1992; Iyamu e Sekgweleo, 2013). Pode ser utilizado na análise e desenvolvimento de sistemas de informação com forte inter-relação entre social e tecnológico de acordo com Iyamu e Sekgweleo (2013), sendo que como indicado por Villiers (2005), é uma técnica adequada para investigação de arquiteturas de conhecimento com integração de áreas científicas, tecnológicas, engenharia, organizacionais e de gestão. Como tal, este método foi utilizado na investigação pelo facto de se abordar temas multidisciplinares em sistemas de informação aplicado à gestão das organizações, além da hipótese de solução obrigar a uma multidisciplinaridade temática ao nível de *business intelligence*, metadados, ontologias e arquitetura empresarial.

3 Estado da arte

Neste capítulo detalha-se os domínios relacionados com a investigação considerando que fazem parte do problema e da hipótese de solução

- **Modelação das organizações (sub-capítulo 3.2):** corresponde ao “Modelo da Organização” da Figura 1.1. Neste sub-capítulo define-se a organização e sua relação com sistemas de informação. Caracteriza-se igualmente o conceito de memória organizacional, bem como a sua relação com *business intelligence* e ontologias. Neste domínio obtemos para a investigação o entendimento dos alicerces da organização ao nível de conceitos (e.g.: objetivos, processos, estruturas orgânicas, clientes, fornecedores, produtos) que são representados enquanto classes com relações e atributos específicos para capturar a estrutura e semântica, como base para se definir as dimensões e métricas de negócio;
- **Gestão de conhecimento (sub-capítulo 3.3):** corresponde à “Representação de conhecimento” na Figura 1.1. Neste sub-capítulo define-se a relação entre dados, informação e conhecimento, como enquadramento para se analisar dos domínios da arquitetura empresarial, bases de dados, metadados, ontologias e *web semântica*. Neste domínio obtemos para a investigação o entendimento de modelos alternativos para se endereçar o problema e hipótese de solução enquanto formas de representação de conhecimento;
- **Sistemas de informação (sub-capítulo 3.4):** corresponde ao “Processamento de informação” e “Exploração de conhecimento” na Figura 1.1. Neste sub-capítulo define-se a arquitetura de sistemas de informação separando os sistemas operacionais dos sistemas informacionais. Sistemas operacionais onde são registados e geridos os eventos de negócio da organização. Sistemas informacionais onde se agrega a informação dos sistemas operacionais dentro do enquadramento do modelo da organização, para se focar na gestão de desempenho. Em adicional, apresenta-se os sistemas de exploração de conhecimento ao nível de *knowledge & data discovery*, *ontology learning*, *data lineage*, processamento de língua natural, *reasoner* e *recommender system*. Tendo por base estes tipos de sistemas, identifica-se o papel de metadados, das dimensões e das métricas como parte da solução;
- **Integração e interoperabilidade (sub-capítulo 3.5):** definição e relação entre integração e interoperabilidade de informação. Neste domínio obtemos para a investigação a base de discussão sobre integração e interoperabilidade face ao tipo de problema em investigação.

3.1 Introdução

As organizações são agrupamentos de pessoas que se estruturam e utilizam recursos técnicos e financeiros em torno de um objetivo comum de acordo com Chiavenato (2005) e Bilhim (2004). Estas organizações, além de terem um ambiente interno composto por processos, pessoas e sistemas de informação, enquadram-se num contexto de ambiente externo onde se devem posicionar face aos clientes, fornecedores, concorrentes, economia e enquadramento legal.

Pela sua dinâmica entre vários componentes, internos e externos, numa lógica contínua de entrada, processamento e saída de informação, as organizações podem ser vistas numa perspetiva sistémica para melhor entendimento da importância da informação e regras de negócio. Para tal, pode-se criar modelos da organização, onde os componentes são representados por conceitos (estrutura) e as regras de negócio são representadas como relações entre conceitos (semântica).

Um dos recursos críticos para as organizações são os seus sistemas de informação. Estes sistemas, suportam a oferta e posicionamento da empresa no mercado, na relação com clientes e parceiros, ou permitem automatizar processos de negócio para ganhos de eficiência e eficácia, tendo por base a gestão da informação da empresa. Estes sistemas de informação têm uma arquitetura própria e podem decompor-se em camadas em resposta a necessidades estratégicas, tácitas e operacionais, segundo Laudon e Laudon (1999). Como tal, como parte desta arquitetura de sistemas de informação, fazem parte não só as aplicações centrais de suporte ao negócio no que é designado por sistemas operacionais (e.g. ERP e soluções específicas de acordo com o sector de atividade), mas igualmente os sistemas de informação de gestão, ou informacionais, no que é designado por sistemas de *business intelligence* (Berkeley, 2006; Moss e Altre, 2003). Os sistemas de *business intelligence* podem decompor-se ainda em sistemas de captura (ETL e *data replication*), armazenamento de dados (*Data Stage*, ODS, DW e DM) e sistemas de exploração (*self-service, reporting, dashboard* e *analytics*). Nesta tese, consideramos os sistemas de captura como informacionais, os sistemas de armazenamento como representação de conhecimento e os sistemas de exploração de informação como sendo exploração de conhecimento.

Os sistemas de informação existem assim num contexto organizacional, sendo que o seu desenvolvimento com suporte a bases de dados não se traduz somente na implementação de novos meios tecnológicos, mas igualmente considerando a arquitetura de informação, segundo Caldeira (2009). Esta arquitetura de informação considera as estruturas de dados sobre as transações de negócio (reflexo operacional da atividade da empresa), agregação das transações para análise e tomada de decisão (reflexo tático e estratégico da empresa), mas igualmente sobre a estrutura da empresa (componentes, vistos como recursos, que devem ser otimizados para fazer negócio), considerando a sua estrutura e a sua semântica. São estes últimos, que permitem representar, analisar e monitorizar a eficácia e eficiência tendo por base os ativos estruturais onde as atividades se realizam e como se realizam.

Mas para além do tratamento da informação, estes sistemas são igualmente essenciais para o conhecimento sobre a própria informação e sobre o seu ciclo de vida (gestão de conhecimento). Para tal, os sistemas recorrem a componentes de modelação, sendo que existem tipos de sistemas específicos só para modelação. Estes sistemas de modelação fazem parte da representação de conhecimento, sendo que a informação sobre as estruturas de dados da empresa está integrada no conceito de metadados, com visões diferentes em vários tipos de sistemas:

- ***Business intelligence***: os metadados são utilizados para guiar e inferir sobre estruturas de dados, o que é fundamental para os processos de ETL, mas igualmente em sistemas modernos de *business visualization* para relatórios e *dashboard*;

- **Arquitetura empresarial:** a base de dados onde estão instanciados os conceitos, propriedades e relações correspondem neste caso a metadados sobre conceitos de negócio, estrutura organizacional, processos, objetivos e componentes tecnológicos;
- **Bases de dados:** a linguagem DDL permite definir os metadados de base de dados. Isto porque uma estrutura de “CREATE TABLE” inclui a definição de tabelas, campos, chaves primárias e chaves estrangeiras (para obter a relação com outras tabelas);
- **Sistemas operacionais:** são normalmente capturados via linguagens de modelação como UML, ficando integrados em regras de negócio específicas de cada sistema ao nível de motores de *workflow*, sistemas de *rule engine*, lógica de programação ou estruturas internas aos programas.

A heterogeneidade dos sistemas utilizados pelas empresas, associados às estruturas de dados estruturados e não estruturados que cada sistema utiliza, pressiona a necessidade de encontrar soluções flexíveis para consolidar numa visão única a informação sobre os próprios dados, ou seja, metadados.

Como resposta a estes desafios, o conceito de metadados tem sofrido evoluções, com particular destaque nas seguintes linhas de investigação:

- Metamodelos como é o caso IMM (2008) e DCMI, mas que exige uma adaptação por todos os sistemas;
- Investigações específicas de alinhamento com arquitetura empresarial no sentido de criar metadados corporativos, como é o caso de Vetterli et al. (2000) e Turco (2010), apesar de não ter uma relação direta com *business intelligence*;
- Investigações específicas de alinhamento com ontologias como é o caso de Scilia (2006), aproveitando a evolução e maturidade em especificações no domínio da web semântica, como é o caso do OWL (*Ontology Web Language*), baseadas em linguagem XML (*Extensible Markup Language*), com reutilização de representação de conhecimento, inferência, agentes inteligentes e processamento de língua natural na interação com utilizador;
- Implementação de bases de dados como ontologia, criando o conceito de bases de dados semânticas como GraphDB (GraphDB, 2016) e investigações específicas de conversão de bases de dados relacionais em OWL, sendo, no entanto, uma preocupação sobre as instâncias de dados, mas que pode ser aplicado para armazenamento de metadados.

Face a esta heterogeneidade entre sistemas e modelos de representação de conhecimento, existe um problema natural de interoperabilidade não resolvido. Por outro lado, de forma mais crítica, existe um problema na relação entre modelo da empresa e sua reutilização para avaliação do desempenho, com base no modelo conceptualizado, através das dimensões e métricas. O que leva ao problema em investigação relativo ao papel da arquitetura empresarial e *business intelligence* na representação e avaliação do desempenho organizacional a partir de um modelo comum, considerando que este modelo comum corresponde a uma ontologia.

3.2 Modelação da organização

3.2.1 Organização

Segundo Chiavenato (2005) uma organização é um conjunto de pessoas que atuam juntas com uma criteriosa divisão do trabalho para alcançar um propósito comum, sendo que estas organizações são um sistema cooperativo racional que constituem a alavanca do desenvolvimento económico e social. Bilhim (2004) separa a noção de organização enquanto entidade social, da noção de organização como referência de condutas e processos sociais relacionados com o ato de organizar atividades, dispor meios relativamente a determinados fins e integrar diversos membros numa determinada unidade.

A organização pode ser assim vista como um conjunto de pessoas organizadas em função de um determinado objetivo, dispondo de recursos técnicos e financeiros, e com uma divisão de tarefas sob a forma de processos. Daí que esta visão de organização pode ser vista como um sistema complexo de processos sobre os quais se pode fazer engenharia para atingir objetivos organizacionais específicos, tendo como base a sua natureza dinâmica, de acordo com Liles et al. (1995). Na definição da organização é várias vezes percecionado o conceito de sistema e dinamismo, quer no contexto interno quer no contexto externo onde se insere, dentro de um sistema mais alargado aonde a organização se enquadra. Esta visão sistémica está alinhada pelo pensamento do biólogo austriaco Ludwig Von Bertalanffy que na década de 50 e 60 definiu a teoria geral de sistemas para compreender o todo a partir de uma análise global de todos os componentes e da interação entre estes. O sistema é assim visto como um conjunto de componentes interligados que operam juntos com um determinado propósito, num determinado ambiente, com o fim de alcançar objetivos comuns e com capacidade de autocontrolo, de acordo com Gomez e Rivas (1989). Esta visão, permite analisar a organização num contexto de engenharia, tendo por base uma representação da mesma através de modelo organizacional, que serve de base à definição e gestão de estratégia, corporizada através de processos que utilizam recursos para atingir objetivos.

Uma vez constituída uma organização, torna-se necessário definir funções que permitem conduzi-la em direção aos seus objetivos. Estas funções podem ser agrupadas por tipos como é o caso do planeamento (determinar o que deve ser feito e como fazê-lo), organização (desenhar e monitorizar a estrutura da organização, as políticas internas e os processos), direção (motivação, liderança e comunicação) e controlo (meios de garantia da eficiência e eficácia organizacional), como referido por Caiado e Caiado (2006). Em função de áreas de negócio específicas, estas funções podem ser aplicadas a gestão comercial, marketing, financeira, organizacional, sistemas de informação, entre outras, de acordo com o sector de atividade da organização. Temos assim que a organização é um conjunto de pessoas organizadas (modelo orgânico) em função de objetivos (modelo estratégico) e funcionando com base em processos (modelo de processos) de acordo com o negócio da organização (modelo de negócio), utilizando informação (modelo de sistemas de informação). Por outro lado, para se poder posicionar no contexto onde se insere, delinear os seus objetivos, criar uma estrutura de comunicação e partilha destes objetivos e aferir sobre a prossecução destes objetivos, as organizações necessitam de delinear o seu modelo de negócio, a sua estratégia e a sua operacionalização. Estes conceitos são abordados nos sub-capítulos seguintes.

3.2.2 Estratégia e modelo da organização

De acordo com Teece (2010), uma organização utiliza de forma implícita ou explícita um modelo de negócio que corresponde a uma arquitetura da mesma que reflete uma hipótese do ponto de vista da organização acerca do que os clientes necessitam, como consomem o que necessitam e de que forma deve estar organizada para responder a essas necessidades, criando assim valor traduzido em rentabilidade.

Por outro lado, para Santos et al. (2002), a estratégia, pode ser vista enquanto forma como uma organização se estrutura, adapta ao meio envolvente e se posiciona no mercado com premissas de sobrevivência na adaptação ao meio envolvente e proteção de vantagens competitivas. Daí a necessidade de utilizar modelos de análise de forças, fraquezas, oportunidades e ameaças (designado por SWOT do termo em inglês de *Strength, Weakness, Opportunities and Threats*) e análise de vantagens competitivas por modelos como das 5 forças de Porter (1985). Esta mesma linha de pensamento está presente em Osterwalder e Pigneur (2010) para os quais um modelo de negócio descreve a forma como uma organização se posiciona no mercado onde se insere para criar as condições para fazer negócio com objetivos de lucros.

A estratégia, formaliza assim, objetivos a médio e longo prazo, fazendo opções com base num modelo de negócio enquanto forma de criar valor focado nos clientes que serve, recursos que utiliza e contexto onde se insere a sua atividade em torno de um modelo que representa a organização, o seu contexto interno e o seu contexto externo. Para tal, a organização é estruturada sob a forma de um modelo de negócio que é formalizado com opções estratégicas para a organização na forma como se estrutura, no posicionamento no contexto de mercado e nos objetivos de valor definidos a médio e longo prazo.

No entanto, não existem consensos sobre a definição de modelo de negócio. Para Hamel (2002), pode incluir a estratégia, enquanto que para Shafer et al. (2005), o modelo de negócio não é uma estratégia, mas pode ser utilizado para analisar e comunicar as escolhas estratégicas. De acordo com vários autores neste domínio temos as seguintes definições:

- Para Morris et al. (2005), o modelo de negócio é uma representação de inter-relações de decisões sobre estratégia, risco, arquitetura e economia para criar vantagem competitiva sustentável em mercados definidos pela organização e utilizando componentes como proposta de valor, clientes, processos internos/competências, posicionamento externo, modelo económico e fatores pessoais associados a investidores;
- Johnson et al. (2008) considera que o modelo de negócio envolve a proposta de valor, fórmula de realização de lucros, recursos e processos, para criar valor;
- Osterwalder e Pigneur (2010) e Osterwalder et al. (2005) referem que o modelo de negócio descreve a lógica como uma organização cria, proporciona e obtém valor. Este autor destaca igualmente a separação entre estratégia e processos de negócio, sendo que posiciona o modelo de negócio enquanto elo de ligação entre os dois conceitos. A estratégia define os objetivos. Os processos de negócio correspondem à operacionalização, isto é, como é que a organização atinge os objetivos. Por sua vez, o modelo de negócio define o que oferece de valor aos clientes e os recursos que utiliza (internos ou externos).

Este modelo de negócio, é assim uma representação da organização e da forma como faz negócio no mercado onde se insere, considerando uma descrição da oferta comercial, clientes, processos e recursos para se atingir os objetivos previstos. Osterwalder (2004) refere que as seguintes atividades estão sempre presentes na gestão do modelo de negócio:

- **Compreender:** captura da estrutura e lógica de negócio para criar um entendimento e alinhamento entre todos os interessados na organização, de acordo com as suas perspetivas e responsabilidades;
- **Analisar:** observar e medir a organização com base no modelo criado, efetuando inclusive comparações com modelos equivalentes, para se identificar as ações corretivas e evolutivas face aos objetivos previstos;
- **Gerir:** garantir uma gestão estratégica e operacional, nas várias dimensões, tendo por base uma visão comum de estrutura e relações dos componentes da organização;
- **Perspetivar:** ação de planeamento e devida análise de recursos necessários ou existentes, para garantir a exequibilidade desse planeamento, permitindo igualmente posicionar o que atingir com os recursos e o que corrigir na estrutura e relação de recursos existentes.

Pietersma e Van den Berg (2014) consolidaram o que designam por modelos de gestão, e que corresponde a políticas, técnicas, métodos e referenciais para modelar e pensar a organização. Para tal, consideram modelos compostos pelas seguintes categorias funcionais de uma empresa:

- Estratégia de empresa e de negócio (e.g. BSC, SWOT, Matriz BCG);
- Organização e governança (e.g. Benchmarking, cadeia de valor);
- Desempenho financeiro (e.g. ABC, EVA);
- Marketing e vendas (e.g. 4P, análise de rede social);
- Operações, cadeias de fornecimento e compras (e.g. Seis Sigma, Lean, Keizen);
- Inovação e gestão da tecnologia para alinhamento com a estratégia da empresa (e.g. Togaf, Itil);
- Recursos humanos e gestão da mudança (e.g. Belbin, ciclo de Deming);
- Liderança (e.g. dimensões culturais, base da pirâmide).

Os modelos considerados por Pietersma e Van den Berg (2014) são compostos por técnicas de modelação e referenciais para serem ajustados por cada empresa, mas que mantêm sempre uma estrutura de conceitos e relações em cada domínio de aplicação. Com o mesmo objetivo, Orofino (2011) consolida vários componentes que fazem parte do modelo de negócio, tal como apresentado na Figura 3.1. Este esquema de componentes considera uma perspetiva financeira na captura de valor, as componentes estratégicas na configuração de valor, os recursos ao nível de cadeia de valor e uma conjugação de oferta comercial e relação com clientes ao nível da criação de valor.



Figura 3.1: Estrutura de modelo de negócio (Orofino, 2011)

Um dos modelos analisados por Orofino (2011) é visto como uma ontologia organizacional segundo Osterwalder (2004) e Osterwalder e Pigneur (2010), sendo que é composto por um conjunto de componentes interrelacionados, como apresentado na Figura 3.2. Neste modelo, destacam-se as atividades enquanto processos, os recursos técnicos, as parcerias e a proposta de valor enquanto oferta comercial a partir do qual se pode definir uma relação com clientes assente em determinados canais para atingir determinados segmentos de clientes definidos estrategicamente para a organização. Na base do modelo de negócio considera-se o equilíbrio entre custos e proveitos (fluxos de rendimentos). A sua utilização é normalmente mais descritiva, motivo pelo qual necessita de um nível de detalhe adicional sob a forma de AE.



Figura 3.2: Estrutura de modelo de negócio (Osterwalder e Pigneur, 2010)

Para Cardoso (1998), ao definir uma estratégia, é necessário analisar o ambiente externo geral (política, sociedade, fiscalidade dos locais geográficos onde a organização opera), ambiente externo específico (mercados, concorrentes, fornecedores e clientes relacionados com os seus produtos e serviços) e ambiente interno (processos, pessoas, tecnologia, estrutura). Temos assim que do ambiente externo determinam-se as oportunidades e ameaças. Do ambiente interno analisam-se as forças e fraquezas, tal como referido por Guimaraes (2006) enquanto proposta de modelo consolidado, apresentado na Figura 3.3.

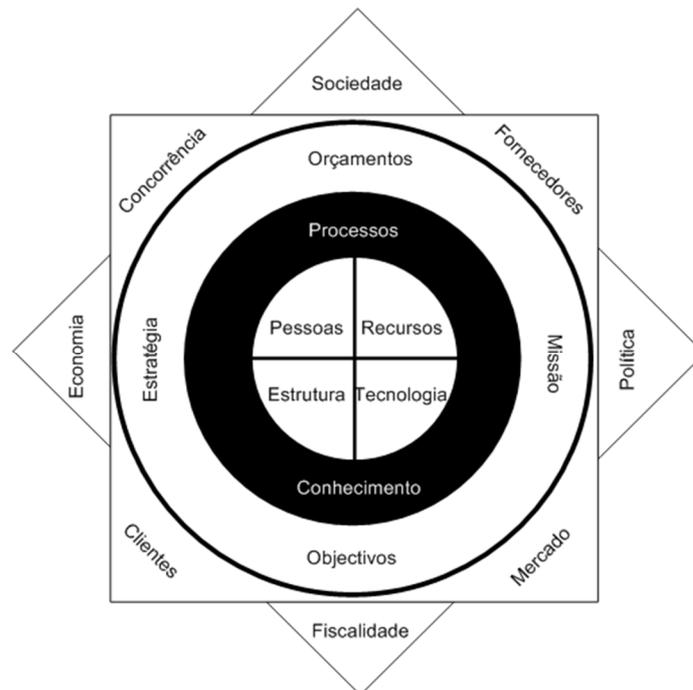


Figura 3.3: Modelo de estratégia (Guimaraes, 2006)

3.2.3 Operacionalização organizacional

Tendo por base a estratégia representada no seu modelo de negócio, as organizações definem a sua operacionalização através de processos e políticas subjacentes aos mesmos. O modelo de processos descreve como devem ser efetuadas as atividades necessárias para a criação de valor envolvendo os recursos necessários e os objetivos a atingir por cada processo, enquanto o modelo de negócio permite definir o que é oferecido (no sentido de criação de valor) de quem para quem, segundo Gordjin et al. (2000).

Um processo de negócio permite agregar de forma sequencial, lógica e racional, como é que se pode de forma eficaz e eficiente atingir um objetivo, decompondo os passos necessários. De acordo com vários autores, a definição de processo considera o seguinte:

- Smith e Fingar (2003) referem que são atividades dinâmicas, coordenadas, colaborativas e transacionais que permitem entregar valor a clientes, internos ou externos;
- Eriksson e Penker (2000) apresentam um conjunto de questões essenciais que um processo tem de ter, como seja o facto ter um objetivo, uma entrada, uma saída, recursos utilizados, ser horizontal à organização, ter um conjunto de atividades sequenciadas e condicionadas, e por fim, ter de criar valor para um cliente interno ou externo.

Um dos aspetos críticos nos processos é assim a capacidade de agregar recursos e criar valor como resposta a um objetivo do processo. Para permitir a medição dos processos, utiliza-se KPI (*Key Performance Indicator*), que correspondem a métricas em *business intelligence*.

3.2.4 Desempenho organizacional

As empresas necessitam de monitorizar se fazem as coisas certas (eficácia) e se fazem as coisas bem (eficiência), sendo que pode ser desperdício fazer bem o que nem se deve sequer fazer de acordo com a estratégia definida, de acordo com Druker (2009). Para o efeito, torna-se crítico equilibrar a eficiência e eficácia efetuando uma aferição sistemática dos resultados e desenvolvimento dos mercados onde as empresas se inserem, de acordo com Kotler e Keller (2006). Para tal, as empresas implementam sistemas de avaliação de desempenho que permitem alinhar as ações de curto prazo de acordo com a estratégia, como referido por Kennerley e Neely (2003).

Este tipo de sistemas corresponde a um conjunto de métricas que permitem quantificar a eficiência e eficácia das atividades segundo Neely et al. (2005) e desta forma medir o alinhamento entre a operacionalização e a estratégia. Como tal, Neely (2002) refere que é um instrumento fundamental para o processo de decisão pois permite recolher dados, consolidar, analisar e interpretar os mesmos, antes da publicação para os vários interessados. Neely et al. (2005) sistematiza assim a ideia, indicando que os sistemas de avaliação de desempenho permitem medir a eficiência e eficácia enquanto suporte ao processo de decisão utilizando um conjunto de processos específicos:

- **Recolha e consolidação de dados:** obter factos em bruto e consolidá-los num conjunto de informações sobre as atividades da organização;
- **Ordenação:** classificar os factos por categorias relevantes para análise de acordo com as perspetivas da organização;
- **Análise e interpretação:** procurar padrões na informação consolidada e ordenada tentando explicar esses padrões;
- **Disseminação:** comunicar as análises efetuadas de acordo com o impacto para a organização e sua estratégia.

Os sistemas de monitorização de desempenho permitem criar uma relação entre gestão, medição e realização, posicionando-se em torno da estratégia, sendo que têm por objetivo medir não só resultados financeiros mas igualmente ao nível de processos, clientes e desenvolvimento de recursos. É o caso das quatro perspetivas dos modelos BSC (*balance scorecard*) de Kaplan e Norton (2008), que evoluiu para ser um modelo de gestão integrado que permite desenvolver a estratégia, planear utilizando mapas estratégicos, seguido de alinhamento da organização em torno da estratégia, planeamento das operações em conformidade, monitorização com base em informação das métricas previstas e adaptar o planeamento, face às mudanças.

Além do BSC, o DEA (*Data Envelopment Analysis*) é igualmente amplamente utilizado pelas organizações, apesar de ser possível uma simbiose entre os modelos, segundo Carvalho (2011), que detalha da seguinte forma cada um dos modelos:

- **No caso do BSC:** são consideradas quatro perspetivas sendo que cada perspetiva tem uma lógica de objetivos medidos com indicadores (isto é, métricas) que permitem avaliar a eficiência e eficácia. A perspetiva financeira permite medir o retorno do capital. A perspetiva do cliente permite medir a satisfação e retenção do cliente. A perspetiva interna permite medir os processos. A perspetiva de inovação e aprendizagem permite medir o desenvolvimento, satisfação e retenção dos colaboradores;
- **No caso do DEA:** centra-se no conceito de Unidades de Desenvolvimento Homogéneo enquanto organizações para aferir por métodos estatísticos como regras de Pareto, a produtividade considerando a relação entre múltiplas entradas e saídas.

Um exemplo de aplicação de BSC a um sector de atividade específico (Banca) é referido por Russo (2006) que apresenta uma estrutura de objetivos e métricas por perspetivas, na Tabela 3.1:

Perspectivas	Objectivos Estratégicos	Métricas
FINANCEIRA	Incrementar Rendibilidade Melhorar Produtividade	Rendibilidade do Capital Resultado por Colaborador Resultado de Comissões Custos / Resultado
CLIENTES	Cultivar Relação de Proximidade Fidelizar	Acréscimo de Crédito / Gerente Acréscimo de Depósitos / Gerente N.º Médio de Produtos / Cliente N.º Médio de Serviços / Cliente
PROCESSOS INTERNOS	Melhorar Operações Melhorar Capacidade de Resposta Modernizar Gestão de Clientes	N.º de Novas Contas N.º de Propostas por Colaborador Prazo de Resposta N.º de Reuniões N.º de Clientes <i>E-Banking</i>
APRENDIZAGEM & DESENVOLVIMENTO	Desenvolver Competências dos Colaboradores Motivar Colaboradores	N.º de Horas de Formação Prémios por Objectivos N.º de Produtos Sondados Taxa de Erros de Serviço

Tabela 3.1: Mapa estratégico para Banca Comercial (Russo, 2006)

3.2.5 Memória organizacional

A memória organizacional corresponde a informação histórica que é utilizada no presente para a tomada de decisão, tendo por base modelos como apresentado na Figura 3.4, de Walsh e Ungson (1991), onde existem processos e um repositório de informação armazenada com perspetivas individuais, culturais, transformações, estruturas e ecologia. No caso dos processos, é considerado o seguinte:

- **Aquisição da informação:** embutida nos processos de decisão, considerando sempre um problema, uma decisão e o feedback dessa decisão face ao problema;
- **Repositório de informação:** captura o contexto de decisões para indivíduos (problemas resolvidos pelos colaboradores no contexto das suas funções), cultura (curva de aprendizagem dos colaboradores para resolver os problemas), transformações (permitem transformar uma entrada numa saída, de forma manual ou automática), estruturas (papéis e regras aplicadas aos colaboradores) e ecologia (estruturas organizacionais de enquadramento dos colaboradores);
- **Exploração de informação:** sobre decisões passadas, no contexto dos problemas do passado face ao contexto do presente.

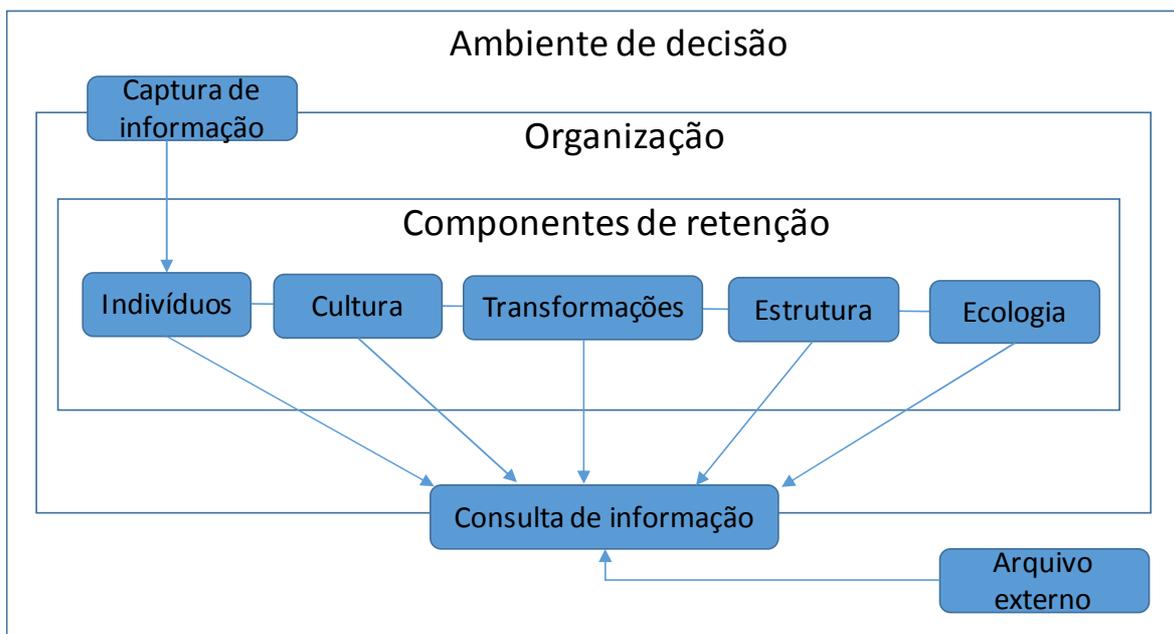


Figura 3.4: Modelo de memória organizacional, adaptado de Walsh e Ungson (1991)

Apesar do modelo ser de 1991, o tema da memória organizacional continua a ser um tema premente, como é o caso de Ochoa et al. (2009) que refere que é necessário ter ferramentas adequadas para aceder à informação considerando aspetos de qualidade de dados ao nível de redundância, contradições e irrelevância de informação armazenada no momento da sua exploração.

3.3 Gestão de conhecimento

3.3.1 Dados, informação e conhecimento

Segundo Rao (2012) a cultura organizacional depende da confiança entre as pessoas que é estimulada por uma gestão que cria um ambiente onde se cria e partilha conhecimento. Este conhecimento, no entanto, está assente em informação tratada pela organização tendo por base dados em bruto gerados ao longo das atividades das organizações.

Dados, informação e conhecimento são assim três conceitos chave, sendo que o conhecimento é visto como um fator de vantagem competitiva nas organizações, pois ao fomentar o desenvolvimento individual e a sua partilha pela organização, melhora o desempenho de toda a organização segundo Davenport e Prusak (1998) e Tiwana (2003). De acordo com Caldeira (2009), a representação de conhecimento pode estar associada ao que denomina como conceito, enquanto representação de um objeto ou acontecimento que entra na composição de um sistema, visto igualmente como convenção social que pode ser representada e transferida para o modelo de dados. Ao ser enquadrado no sistema, permite uma interação e gestão, integrada numa lógica de inferência, tratamento, distribuição e alinhamento contínuo para proteção evolutiva da sua mutação, mantendo coerente o conceito e sua evolução, num contexto de volume enorme de dados, conceitos e relações.

Dados são factos que ocorrem nas organizações sendo que a informação corresponde a uma organização desses dados e por sua vez o conhecimento corresponde à informação organizada de acordo com um determinado sentido. Por definição em dicionário, o conhecimento é o que permite tornar presente aos sentidos um objeto de modo a obter um entendimento ou representação adequado. Mas associado ao conhecimento, podemos ainda relacionar os conceitos de representação (descrever factos, pessoas e ideias através de símbolos), raciocínio (argumentar) e aprendizagem (pensamento para obter conhecimento).

Tiwana (2003) refere que o conhecimento está presente nos documentos, processos, práticas e normas. Para Davenport e Prusak (1998), o conhecimento não é somente dados ou informação, mas sim uma mistura formada por experiências, valores, informação de contexto e perspicácia hábil que proporciona uma estrutura para a avaliação e incorporação de novas experiências e informação, tendo origem e aplicação na mente humana. Nas organizações, costuma estar embebido não só nos documentos ou repositórios, mas também em rotinas, processos, práticas e normas organizacionais. Para Nonaka (1994), existem dois tipos de conhecimento e formas de conversão de conhecimento entre estes tipos, tal como apresentado na Figura 3.5 e que se pode resumir da seguinte forma:

- **Tipo de conhecimento tácito:** reside na mente humana e depende da sua experiência. Corresponde a modelos mentais, esquemas, paradigmas, crenças e pontos de vista que permitem a um indivíduo posicionar-se no mundo;
- **Tipo de conhecimento explícito:** encontra-se em formas externas como documentos e bases de dados. Corresponde a conhecimento objetivo, racional e codificado, permitindo como tal a sua representação em linguagens formais.

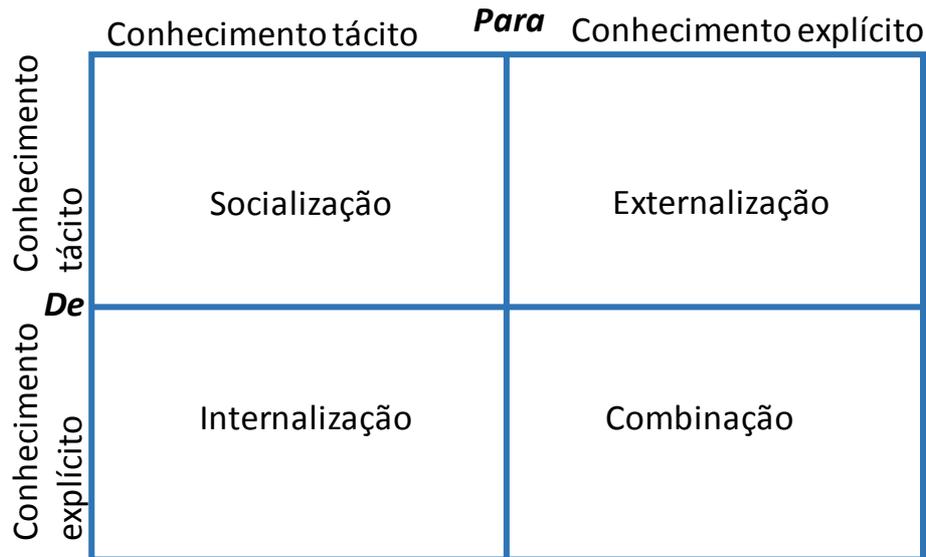


Figura 3.5: Modelos de criação de conhecimento, adaptado de Nonaka (1994).

Sobre as formas de interação ou conversão de conhecimento, Simard e Rice (1996) refere que o conhecimento resultante das interações entre indivíduos, torna-se importante em termos de gestão enquanto recurso, na medida em que pode contribuir como vantagem competitiva e otimização do desempenho da organização. Nonaka (1994) refere que essas formas de interação correspondem a quatro combinações:

- **Socialização:** transformação de conhecimento tácito em tácito, também conhecido como aprendizagem a partir de troca de experiências entre indivíduos ou processos de raciocínio;
- **Externalização:** conversão de conhecimento tácito em explícito, através da representação de conhecimento;
- **Combinação:** composição de conhecimento explícito com outros conhecimentos explícitos, levando normalmente à geração de novos conhecimentos explícitos;
- **Internalização:** conversão de conhecimento explícito em tácito, por aprendizagem de conhecimento explícito disponibilizado, ou como resultado da combinação de conhecimento explícito.

A representação de conhecimento pode ser vista ainda como a aplicação lógica de ontologias no sentido de construir modelos computacionais para determinados domínios permitindo a sua utilização em processos raciocínio, como referido por Sowa (2000a). Neste sentido, o conhecimento é representado por ontologias enquanto descrição de conceitos e relações que existem entre eles, como um vocabulário, o que constitui um universo de discurso representado com algum formalismo declarativo sobre o qual os agentes devem comprometer-se para poderem comunicar nesse universo de discurso, sem terem de lidar com uma teoria geral do conhecimento, de acordo com Gruber (1993). Sowa (2000b), considera ainda que desta forma pode-se então criar modelos computacionais para determinados domínios, mas visto como um tema multidisciplinar que cruza as teorias e técnicas da lógica (estrutura formal e regras de inferência), ontologias (tipo de coisas que existem num determinado domínio) e computação (sem modelos computacionais a lógica e ontologias não podem ser implementadas em programas de computadores).

Para construir estes modelos computacionais, utilizam-se esquemas, que podem ser procedimentos (sequência de instruções em linguagens de programação), regras (bases de dados de expressões em lógica formal de predicados utilizando-se linguagens como o Prolog), redes (grafos com nós como objetos e arcos como relações, utilizando-se redes semânticas e mapas cognitivos) e estruturas (extensão das redes, em que cada nó é uma estrutura designada por *Slot* com valores, como utilizado pelos *frames* e *scripts*), RDF (*Resource Description Framework*) e OWL. Representar o conhecimento é assim construir Ontologias para permitir a partilha de conhecimento, possibilitar a reutilização, tornar o domínio explícito e analisar o domínio de conhecimento onde operamos, como indicado por Noy e McGuinness (2001). Daí a importância da representação de conhecimento, segundo Davis et al. (2010), que refere ainda que permite entender e descrever a riqueza do mundo, podendo ser entendida em função dos cinco papéis principais que desempenha:

- Substituto das coisas, permitindo raciocinar em vez de agir sobre o mundo;
- Conjunto de compromentimentos com ontologias que permitem identificar de que forma é que se pensa sobre o mundo;
- Parte da teoria do raciocínio no sentido em que representa o conceito fundamental do raciocínio, ou seja, o conjunto de inferências que ratifica e o conjunto de inferências que recomenda;
- Meio para a computação eficiente, pois representa a forma como se organiza a informação para facilitar a aplicação das inferências recomendadas;
- Linguagem utilizada para o ser humano se exprimir acerca do mundo.

3.3.2 Gestão de dados e conhecimento

A gestão de conhecimento corresponde à gestão sistemática do conhecimento vital da organização, associada a processos de criar, recolher, organizar, difundir, utilizar e explorar o conhecimento, para converter conhecimento individual em corporativo, segundo Skyrme e Amidon (1998), incluindo o conhecimento não sistematizado e experiências pessoais.

Existem vários modelos de gestão de conhecimento sendo um deles apresentado por Davenport e Prusak (1998) para criação, codificação e a transferência, que detalha o modelo entre geração (criação e aquisição), codificação (refinamento e armazenamento), transferência (transferência e partilha) e utilização. A utilização corresponde à aplicação no contexto da necessidade de cada individuo e organização para suporte à tomada de decisão.

Para Dignum (2003), o conhecimento visto e gerido com estas considerações corporativas, é uma memória organizacional que permite preservar e partilhar conhecimento e experiências na organização, integrando conhecimento disperso. Dignum (2003) acrescenta que normalmente, a implementação deste conceito, de memória organizacional falha, pelo facto de se requerer esforço adicional em vez de fazer parte do processo normal de criação de conhecimento integrado com os processos de negócio. Daí que Dignum (2003) refira que a gestão de conhecimento deve suportar as pessoas no contexto das suas tarefas, em vez de adicionar esforço extra no seu trabalho.

As organizações precisam de reconhecer o conhecimento como um ativo e geri-lo de forma eficaz para terem sucesso, segundo Boumarafi e Jabnoun (2008). Por outro lado, a gestão de conhecimento depende de sistematização e abordagens que permitam o crescimento da informação e do conhecimento, envolvendo pessoas, fluxos de informação e ferramentas adequadas, tal como referido por Rao (2012).

Por essa razão, existe uma tendência para associar aos processos de negócio modelos de gestão de conhecimento, levando inclusive a conceitos como a gestão de processos orientada à gestão de conhecimento BPOKM (*Business Process Oriented Knowledge Management*) definido por Abecker et al. (2001). Essa mesma razão, leva-nos a considerar a importância da visão do conhecimento, como lógica derivada de transformação de dados em informação e por fim em conhecimento, mas estruturada como uma Ontologia integrada num conceito completo de arquitetura empresarial para garantir uma coesão de representação, utilização e gestão, utilizando-se o *business intelligence* para extrair de forma consistente a informação para tomada de decisão estratégica ou operacional.

Angeloni e Fernandes (2000) na sua definição de organizações do conhecimento enquanto entidades focadas na criação, armazenamento e partilha do conhecimento, define um modelo objetivo na inter-relação entre as organizações e sistemas de informação, destacando no modelo o papel das pessoas, como apresentado na Figura 3.6.

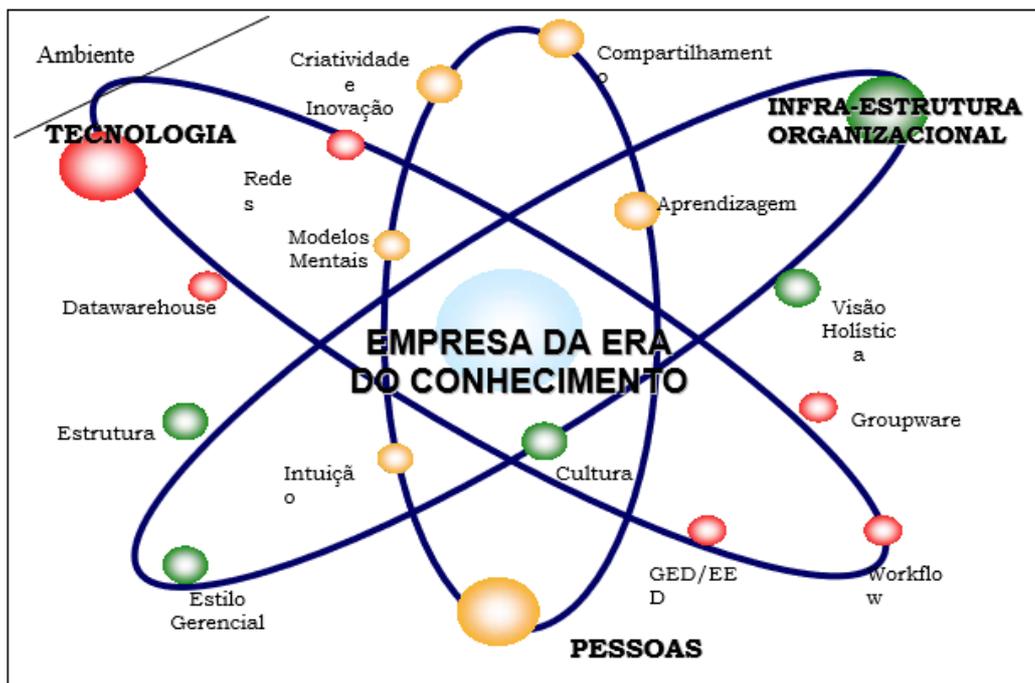


Figura 3.6: Inter-relação de conceitos nas organizações do conhecimento (Angeloni e Fernandes, 2000)

Como evolução da gestão de conhecimento, um dos conceitos emergentes nas organizações é o *data governance*, onde enquadrámos igualmente o conceito de *knowledge & data discovery*, enquanto disciplina de controlo de acesso, gestão, utilização, melhoria, monitorização e proteção de informação organizacional de acordo Khalid (2010). Este conceito define uma abordagem organizacional para a gestão de dados e informação que é formalizada através de políticas e procedimentos adequados ao ciclo de vida dos dados desde que são criados até ao momento em que são utilizados, para se garantir requisitos de confidencialidade, integridade, acessibilidade, disponibilidade e qualidade dos dados.

Para a implementação de abordagens de *data governance*, torna-se necessário focar nas seguintes áreas:

- Definição de níveis de *data governance* em termos de papéis e responsabilidades dentro da orgânica definida;
- Gestão de segurança e risco tendo por base a análise de dados sensíveis;
- Garantir a adequada gestão de inventário e conteúdo de dados devidamente atualizado com definição sobre que dados são necessários e existem, para que são utilizados, como estão estruturados, onde residem e qual o seu nível de sensibilidade/risco;
- Definição de estratégias para garantir a qualidade dos dados de forma preventiva e reativa, estabelecendo os mecanismos para o efeito, incluindo para a sua correção;
- Garantir que a partilha e distribuição de dados obedece aos princípios de segurança e risco definidos.

O conceito de *data governance* integrado em sistemas, políticas e processos internos nas organizações permite entregar a informação que os utilizadores necessitam e de acordo com nível de autorização definido para acesso a essa informação. Pela necessidade de garantir um inventário e conteúdos de dados devidamente classificados, complementado com monitorização dos dados ao longo do seu ciclo de vida, torna-se um conceito transversal aos sistemas de informação, integrando igualmente domínios como gestão de metadados e gestão de conhecimento de forma alargada. Neste contexto, alguns fabricantes de soluções de sistemas de informação enquadram vários componentes da sua oferta para uma adequada gestão de dados organizacionais. Tal é o caso do modelo IBM *Data Governance Council Maturity Model* [IBM] (2007), referido igualmente por Davis (2010), apresentado na Figura 3.7, onde para além do modelo de maturidade, destaca-se no contexto da nossa investigação, as disciplinas de suporte onde se enquadra a “*data architecture*” e “*classification & metadata*”, além das disciplinas principais onde se destaca “*information life cycle management*”.

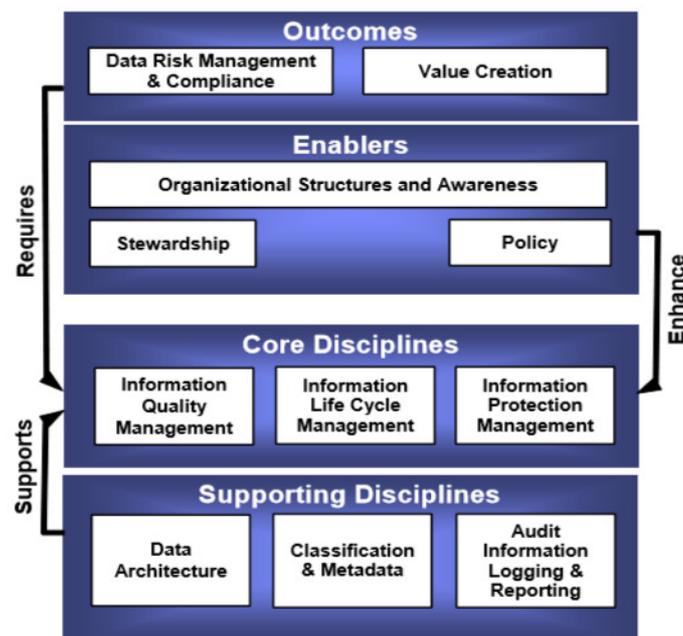


Figura 3.7: IBM Information Governance Maturity Model (IBM, 2007; Davis, 2010)

3.3.3 Representação de conhecimento

3.3.3.1 Arquitetura empresarial

Rodrigues (2014) na sua investigação sobre "Fatores determinantes do valor das arquiteturas empresariais", indica que a arquitetura empresarial assume um papel importante nas organizações devido à necessidade de alinhamento contínuo entre sistemas de informação e negócio, pelo papel da arquitetura na gestão da mudança, por permitir a redução de custos e suportar a redução da complexidade. Qualquer dos motivos depende da existência de um modelo da organização, que permita um pensamento contínuo com base num desenho de uma realidade da organização, vista de forma holística. Este modelo da organização corresponde assim a uma arquitetura empresarial, sendo que na literatura sobre este domínio, existem várias definições alternativas:

- Para o *Institute of Electrical and Electronics Engineers [IEEE] (2000)*, que corresponde à norma ISO/IEC 42010:2007, pode ser definido como: *“The fundamental organization of a system, embodied in its components, their relationships to each other and the environment, and the principles governing its design and evolution”*;
- Lankhorts (2006) refere que a arquitetura empresarial é um conjunto de princípios, métodos e modelos utilizados no desenho e concretização de uma estrutura organizacional empresarial, processos de negócio, sistemas de informação e infraestrutura. Neste caso, coloca-se a ênfase nos componentes críticos do modelo de negócio, assentes numa lógica de componentes e relações derivados do conceito de arquitetura, que é utilizado igualmente na conceção de modelos de negócio. Como tal, é uma forma de colocar sob a forma de arquitetura o modelo de negócio;
- Microsoft (2009) considera que arquitetura empresarial como: *“Defining the overall form and function of systems (business and IT) across an enterprise (including partners and organizations forming the extended enterprise), and providing a framework, standards and guidelines for project-level architectures. The vision provided by the Enterprise Architecture allows the development of consistent and appropriate systems across the enterprise with the ability to work together, collaborate, or integrate where and when required.”*;
- Group (2013) descreve a arquitetura como sendo *“Architecture descriptions are formal descriptions of a system, organized in a way that supports reasoning about the structural and behavioral properties of the system and its evolution. They define the components or building blocks that make up the overall system and provide a plan from which products can be procured, and subsystems developed, that will work together to implement the overall system.”*. O que é uma caracterização equivalente à de ontologias ao referir que corresponde a uma descrição formal de um sistema, que suporta inferência com base na estrutura (classes) e comportamento (propriedades) do sistema.

Na sua análise sobre arquitetura empresarial, Rodrigues (2014) refere que "uma arquitetura empresarial consiste numa descrição da organização através de uma coleção coerente e abrangente de princípios, modelos e outros documentos que facilitam uma visão holística e integrada dos principais componentes e dos seus relacionamentos".

Como parte da sua análise, Rodrigues (2014) destaca os seguintes aspetos críticos a serem considerados numa arquitetura empresarial:

- **Abrangência:** em termos de abrangência, apesar de inicialmente ser mais utilizada em sistemas de informação, torna-se cada vez mais consensual a sua utilização igualmente para componentes de negócio, devidamente relacionados entre si e com os componentes de sistemas de informação;
- **Tipo de componentes:** no caso de negócio, são considerados componentes como a missão, as estratégias, os objetivos, as políticas, os princípios e os processos de negócio. No caso de sistemas de informação, são considerados componentes como as aplicações, as bases de dados, os interfaces, os protocolos e a infraestrutura tecnológica de suporte. Normalmente, estes componentes são organizados em domínios, perspetivas ou arquiteturas, sendo os mais comuns os domínios do negócio, dos dados ou da informação, das aplicações e tecnológico;
- **Horizonte temporal:** neste caso, deve-se considerar a necessidade de desenhar o estado atual e futuro da organização. Isto porque, a arquitetura não deve ser somente um conjunto de descrições da organização, sendo que deve igualmente ser parte de um processo contínuo e iterativo, pois não são as descrições que geram valor, mas sim a sua implementação e utilização.

O conceito de arquitetura empresarial depende do seu enquadramento em termos de recursos sobre os quais se considera a definição de uma estratégia que posiciona esses mesmos recursos com um fim. Como tal, é necessário considerar como se chega ao conceito de arquitetura empresarial a partir de estratégia, que tipos de modelos de arquitetura e processos de criação e gestão existem, para um melhor enquadramento de como é que esta arquitetura devidamente representada, corresponde a uma ontologia da organização e serve de base para um sistema integrado de informação de gestão e sua própria governação. A arquitetura empresarial permite atingir este objetivo ao representar a organização por níveis de detalhe e dimensões, permitindo conhecer os conceitos da organização, a sua estrutura e suas relações, garantindo rastreabilidade e apoio ao desenho dinâmico, isto é, modelação em ação. Towers (2005) refere que a arquitetura empresarial é utilizada com os seguintes objetivos:

- Suporte à implementação da estratégia;
- Clarificação de responsabilidade, tendo por base o conhecimento do posicionamento de cada função, das suas tarefas e da relação entre todos os componentes da organização;
- Definir um *framework* que permita identificar onde estamos e para onde queremos ir, detalhando o caminho e monitorizando o processo de transformação.

Land et al. (2009) posicionam a arquitetura empresarial como forma das empresas gerirem a complexidade e risco. Referem como aplicações práticas, a sua utilização enquanto descrição da situação da empresa, análise de desvio entre situação atual e situação futura desejadas, apoio a planeamento e comunicação da estratégia, apoio a planeamento tático, apoio a planeamento operacional incluindo gestão de carteira de projetos de transformação e apoio à seleção de soluções quer ao nível de processos, orgânica e sistemas de informação.

Em termos de tipologia, Godinez et al. (2010) referem a relação entre vários tipos de arquitetura integradas no conceito de arquitetura empresarial, considerando a arquitetura estratégica (metas e objetivos), arquitetura de negócio (organização e estratégia), arquitetura aplicacional (aplicações informáticas), arquitetura informacional (informação independente do seu suporte) e arquitetura tecnológica (infraestrutura técnica de suporte).

Genericamente, de acordo com *The Open Group Architecture Framework [TOGAF]* (2013), uma arquitetura empresarial deve considerar os seguintes níveis:

- **Arquitetura de negócio:** visão, missão, objetivos, estruturas orgânicas, localizações, processos, oferta comercial, canais de distribuição e segmentação de clientes;
- **Arquitetura de informação:** conceitos de informação e aplicações informáticas para a sua gestão;
- **Arquitetura tecnológica:** infraestrutura tecnológica considerando o suporte às aplicações informáticas, mas igualmente as redes, comunicações e serviços base tecnológicos para garantir a fluidez da comunicação em suporte digital.

Considerando os vários níveis de arquitetura, cada um correspondente a um modelo de uma perspetiva da organização, pode-se então considerar que os vários conceitos da organização e suas relações, podem ser agrupados em arquitetura estratégica, negócio, aplicacional, informacional e tecnológica, tendo por base a tipologia indicada por Godinez et al. (2010), pois é mais desagregada do que a apresentada em TOGAF (2013). No entanto, pode-se isolar a arquitetura funcional, ou também designada por processos, que é mencionada como parte da arquitetura de negócio por TOGAF (2013).

Como forma de representação, existem vários *frameworks* com vários modelos de representação das dimensões e seu detalhe. Algumas delas são específicas para instituições governamentais como é o caso do DoDAF (*Department of Defense Architecture Framework, EUA*), e-GIF (*eGovernment Interoperability Framework, Reino Unido*), E2AF (*Extended Enterprise Architecture Framework of Institute for Enterprise Architecture Development*), FEAF (*Federal Enterprise Architecture Framework, EUA*), MODAF (*Ministry of Defense Architecture Framework*) e NAF (*NATO Architecture Framework*). Além destes *frameworks* governamentais, existem outras para utilização empresarial, sendo de destacar as seguintes:

- **Zachman:** criado em 1987 por Zachman (2009), é um modelo genérico que segundo o autor pode ser visto como uma ontologia organizacional composta por um conjunto de componentes que definem um objeto de modelação. Pode ser aplicado a um sistema de informação a implementar ou a uma organização;
- **MEAF (*Metis Enterprise Architecture Framework*):** criado pela empresa Troux Technologies, atualmente integrada na PlanView (PlanView, 2017), a partir de modelos de várias empresas e que está na base da sua ferramenta de arquitetura empresarial, *IT portfolio management, IT governance management e metadata management*;
- **TOGAF ArchiMate (*The Open Group Enterprise Framework*):** apresenta um modelo de arquitetura de negócio, aplicações e tecnologia, uma metodologia para a sua implementação e uma linguagem de notação que pode ser reutilizada por várias ferramentas, e que é designada por ArchiMate, publicado pela The Open Group (GROUP, 2012, 2016).

No caso do modelo de Zachman (Zachman, 2009) apresentado na Figura 3.8, pode-se identificar a necessidade no domínio de sistemas de informação para relacionar de forma vertical em termos de níveis de detalhe os dados, os processos e programas informáticos, mas igualmente as pessoas, distribuição geográfica e mesmo motivações e eventos críticos. Estes dois últimos mais associados aos objetivos estratégicos. No modelo de Zachman considera-se uma estrutura matricial com duas dimensões:

- Para definir os dados (*what*), as funções (*how*), a localização geográfica (*where*), as pessoas (*who*), tempo (*when*) e a motivação (*why*) associados aos artefactos;
- Para definir o nível de detalhe em termos de âmbito (*scope*), modelo de negócio (*business model*), modelo de sistemas (*system model*), modelo de tecnologias (*technology model*) e representação detalhada (*component assemblies e operations instance classes*).

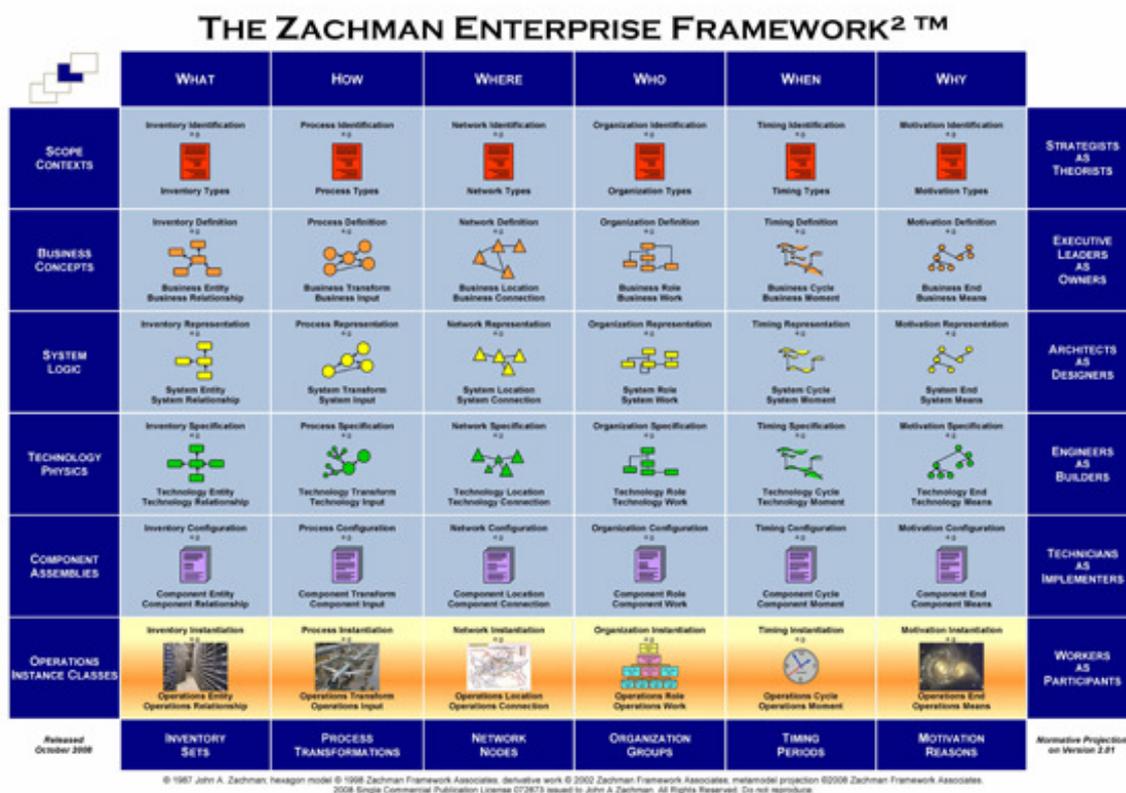


Figura 3.8: Framework de Zachman (Zachman, 2009)

Pela importância de definir uma metodologia, tipologia de arquiteturas e linguagem de notação, o ArchiMate (GROUP, 2012, 2016; TOGAF, 2013) apresentado na Figura 3.9 considera uma metodologia onde se destaca a arquitetura de negócio, a arquitetura aplicacional e a arquitetura tecnológica. Cada um destes níveis de arquitetura tem perspectivas de estrutura, informação e comportamento. Na combinação entre tipo de arquitetura e perspectivas, existem componentes que são utilizados para modelar a organização.

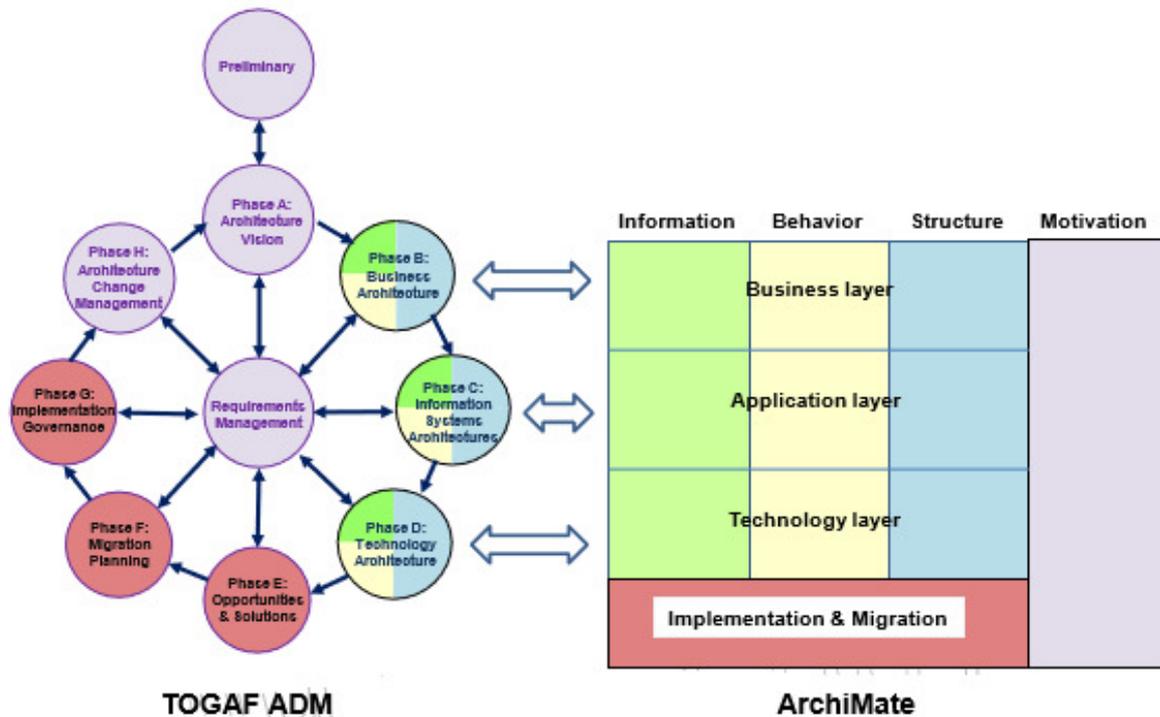


Figura 3.9: Modelo de arquitetura TOGAF e ArchiMate (GROUP, 2012)

Cada tipo de componente-tem um código de cores, sendo que existem relações entre os componentes de várias arquiteturas, tal como apresentado no modelo de relacionamento da arquitetura ArchiMate (GROUP, 2012) na Figura 3.10.

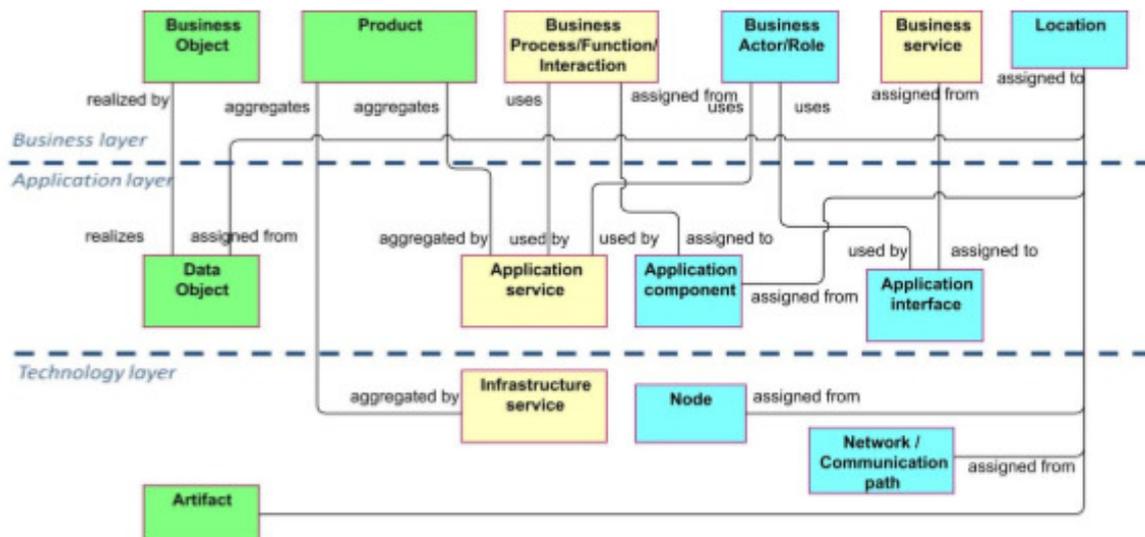


Figura 3.10: Relação entre níveis de arquitetura ArchiMate (GROUP, 2012)

O *business layer* no modelo ArchiMate (GROUP, 2012) considera a relação entre produtos enquanto proposta de valor e os processos necessário para concretizar esse valor ao nível da cadeia de valor que envolve colaboradores com determinações funções e perfis, além de outros recursos, como apresentado na Figura 3.11.

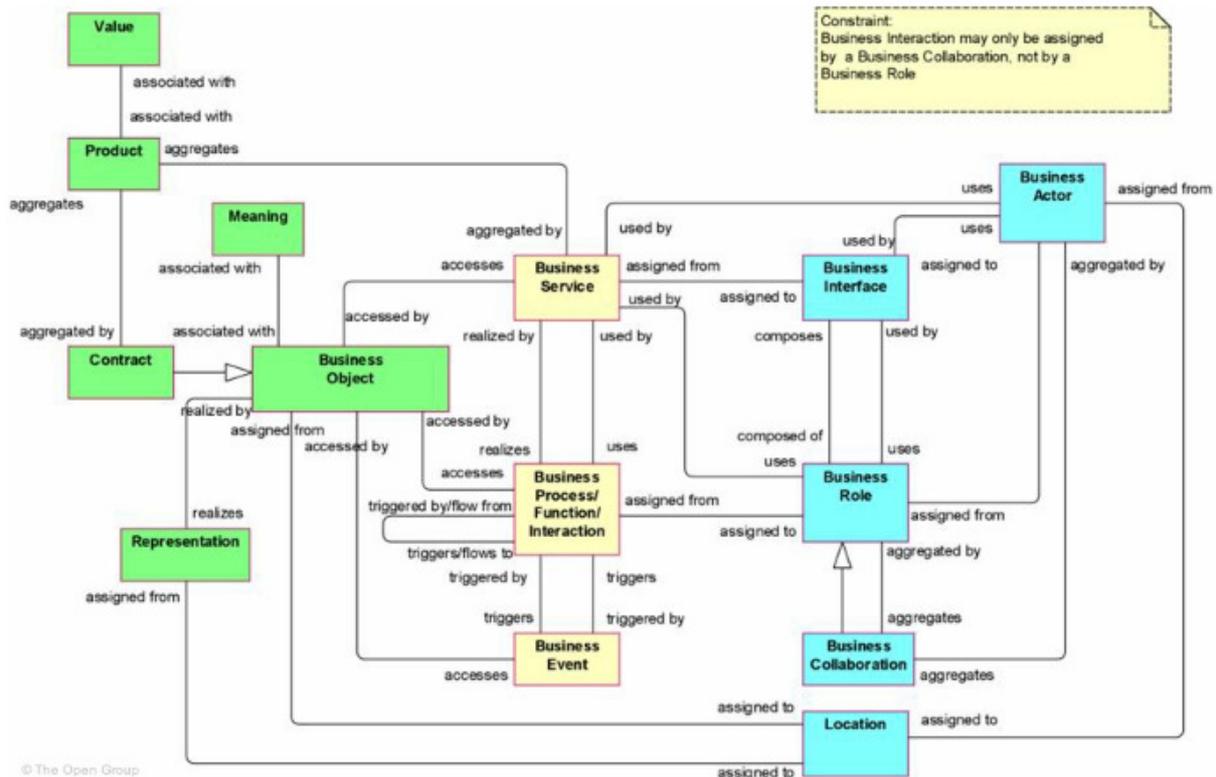


Figura 3.11: Business Layer ArchiMate (GROUP, 2012)

O *application layer* no modelo ArchiMate (GROUP, 2012) considera a relação entre *data object* e *application service*, isto é, entre objetos de dados e serviços aplicacionais que os criam e utilizam, além de outros componentes, como apresentado na Figura 3.12.

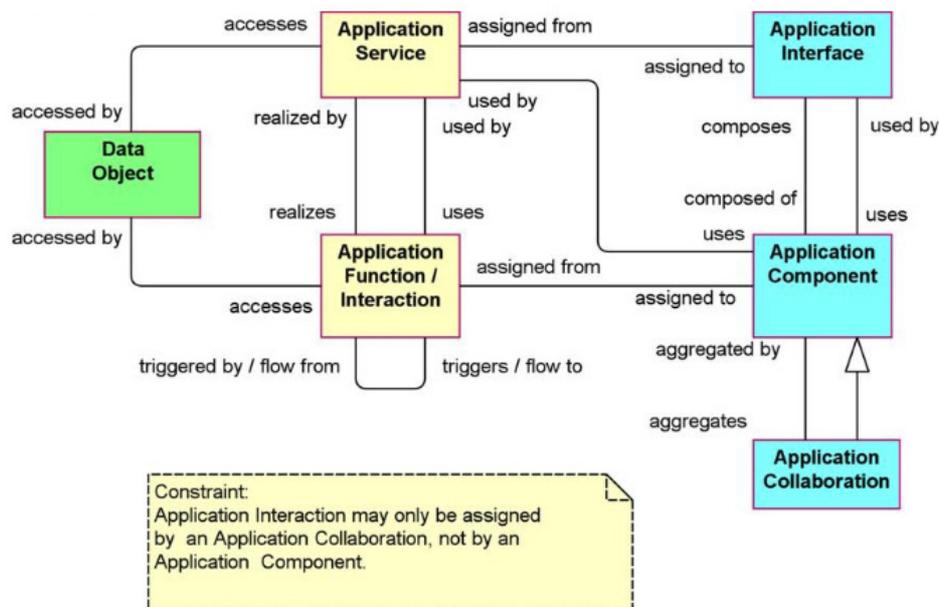


Figura 3.12: Application Layer ArchiMate (GROUP, 2012)

O *technology layer* no modelo ArchiMate (GROUP, 2012) considera a relação *artifact* enquanto conceito que necessita de infraestruturas de *hardware* e *software* para a sua concretização, como é o caso de *system software*, *devices*, *network* e *infrastructure service*, entre produtos componentes, como apresentado na Figura 3.13.

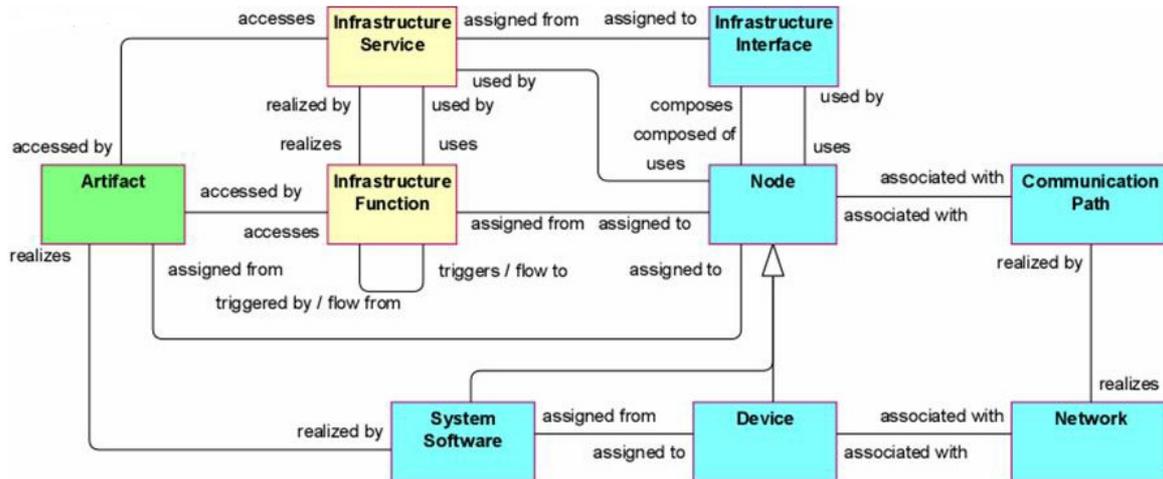


Figura 3.13: Technology Layer ArchiMate (GROUP, 2012)

Cada *framework* normalmente tem a sua própria metodologia de implementação que permite definir os requisitos, as atividades críticas, a monitorização pós implementação, além de considerações sobre intervenientes no processo de construção e gestão da arquitetura.

No caso do *framework* de Zachman existem várias abordagens, sendo que se deve destacar o método definido pela OMG (*Object Management Group*), Frankel et al. (2003) e Ostadzadeh et al. (2003), tendo por base a norma MDA (*Model Driven Architecture*) que permite não só a construção da arquitetura (exceto a perspetiva planeamento, perspetiva de organização e a dimensão “quem” na perspetiva de construção e subcontratação) como também a criação de código a partir dos modelos, considerando três fases:

- **CIM (*Computation Independent Model*)**: modelos e vocabulários de domínio que representam um sistema de forma conceptual, e como tal, independente de uma visão computacional;
- **PIM (*Platform Independent Model*)**: modelos lógicos ainda numa vista independente da plataforma, mas já com uma perspetiva computacional;
- **PSM (*Platform Specific Model*)**: modelos físicos já com numa dependência da plataforma computacional.

De acordo com Gartner (2015a) existem várias ferramentas comerciais neste domínio. Todas as ferramentas têm o mesmo tipo de capacidades em termos de modelação da empresa, incluindo processos de negócio, e permitindo explorar o modelo para gestão estratégica e gestão de IT. Permitem inclusive definir metamodelos novos, ou optar por alguns mais conhecidos. Considerando ferramentas comerciais e *open source*, podem-se destacar as seguintes o ARIS, Enterprise Architect, IBM Rational System Architect, PlanView, Mega e Archi.

Rajabi et al. (2013) consideram que um dos problemas na utilização de uma arquitetura empresarial tem a ver com o facto de os dados utilizados para a sua modelação não estão baseados numa base de conceitos de dados da organização. Para este autor, este problema pode levar problemas de inconsistência e interoperabilidade da arquitetura de informação.

Partindo do princípio de criação de uma ontologia empresarial que seja possível de relacionar com o *framework* de Zachman (Zachman, 2009), Rajabi et al. (2013) propõe o modelo apresentado na Figura 3.14. No entanto, observando os conceitos modelados pelo autor em UML, correspondem em grande medida aos conceitos presentes em ArchiMate. Como tal, ressalva-se desta abordagem a visão do problema de interoperabilidade abordado nesta tese e a visão de uma arquitetura empresarial como ontologia empresarial, baseada em conceito de dados, que é igualmente abordado na nossa investigação.

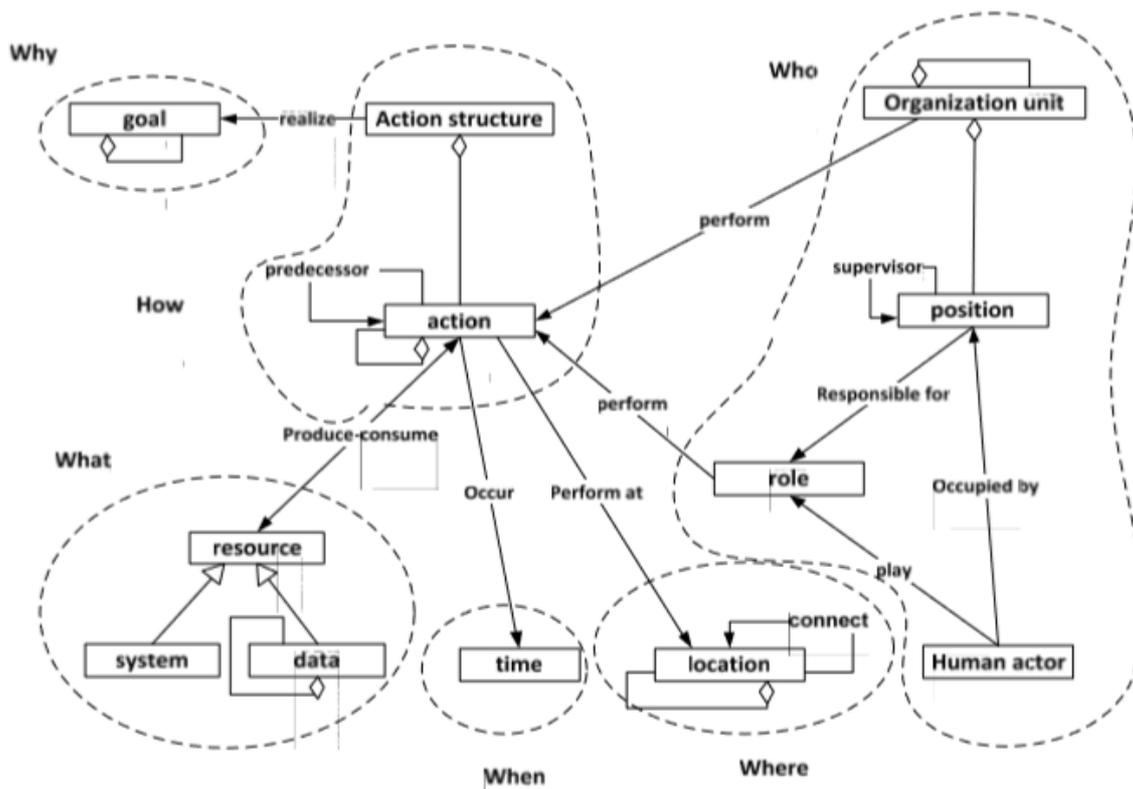


Figura 3.14: Framework de Zachman (Zachman, 2009)

Para a nossa investigação, iremos considerar uma arquitetura empresarial de acordo com o modelo na Figura 3.15, que tem por base a análise efetuada sobre as várias componentes que fazem parte da modelação das organizações (sub-capítulo 3.2) e nos referenciais de arquitetura empresarial analisados neste sub-capítulo, utilizando a linguagem de notação ArchiMate. Este referencial considerar os seguintes níveis que agregam estruturas e semântica de conceitos chave de uma organização que devem ter uma correlação com dados, funcionando assim como um modelo para criar uma base de alinhamento entre sistemas de informação e negócio:

- **Arquitetura estratégica:** definição da visão, missão, princípios e políticas gerais da organização;
- **Arquitetura organizacional:** estruturas orgânicas e localizações;
- **Arquitetura de negócio:** definição da oferta de produtos e serviços, canais de distribuição e segmentação de clientes;

- **Arquitetura funcional:** modelo de funcionamento detalhado em processos, atividades e tarefas para atingir os objetivos da organização;
- **Arquitetura informacional:** estruturas de informação necessária para os processos;
- **Arquitetura aplicacional:** aplicações informáticas, em termos de serviços de suporte ou que permitem integrar e inovar no modelo de prestação de serviços ou integrados em produtos comercializados;
- **Arquitetura tecnológica:** infraestrutura tecnológica considerando o suporte às aplicações informáticas, mas igualmente as redes, comunicações e serviços base tecnológicos para garantir a fluidez da comunicação em suporte digital.

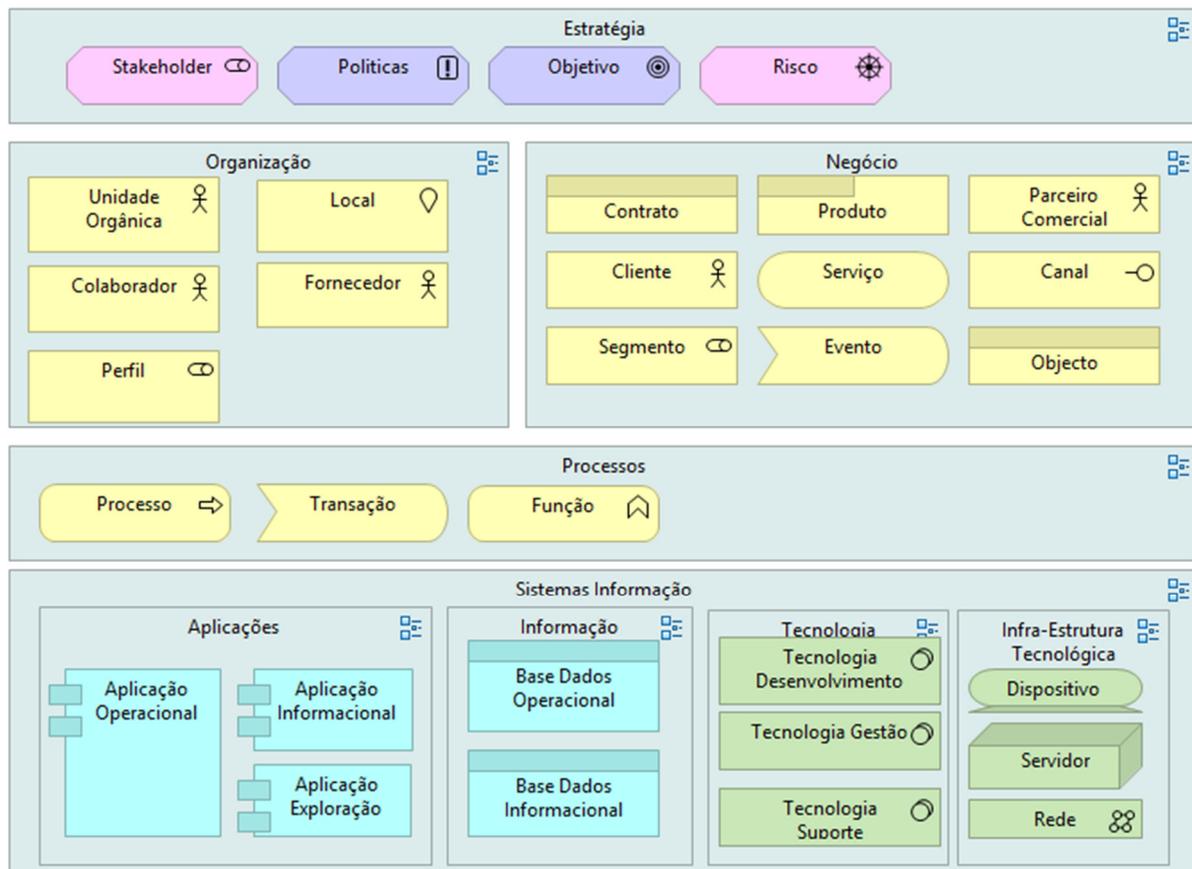


Figura 3.15: Modelo de arquitetura empresarial

3.3.3.2 Metadados

Metadados é um conceito de se obter e gerir dados sobre dados, visto como descrição de informação sobre as próprias de estruturas de dados numa organização e suas aplicações informáticas ou outros ativos tecnológicos. Metadados corresponde assim a informação estruturada que descreve, explica e localiza os recursos de informação. Por essa razão, autores como NISO (2004) definem metadados como *“Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called ‘data about data’ or ‘information about information’”*.

De acordo com outros autores, os metadados podem ser vistos da seguinte forma:

- Marco (2013) define metadados como qualquer dado físico guardado em aplicações informáticas e conhecimento existente em aplicações ou em colaboradores, que representam aspetos internos ou externos à organização, incluindo informação sobre processos de negócio, regras e estruturas de dados. É o conhecimento existente em aplicações ou em colaboradores, que representam aspetos internos ou externos à organização, incluindo informação sobre processos de negócio, regras e estruturas de dados;
- Tozer (1999) considera o conceito de metadados como algo relacionado com o comportamento e descrição de outros dados;
- Inmon et al. (2010) no contexto do que definem como Data Warehouse 2.0, consideram metadados como um dos componentes fundamentais ao referir que *“metadata is one of the cornerstones of the DW 2.0 data warehouse. Metadata is the key to reusability of data and analysis. Metadata enables the analyst to determine what has already been built. Without metadata, the analyst has a hard time finding out what infrastructure is already in place. There are four levels of metadata that includes enterprise, local, business, and technical”*;
- Além de definições de metadados relacionadas com a estrutura de informação sobre os dados existentes, para permitir gerir e partilhar a informação, é igualmente relevante considerar-se nos metadados a perspetiva semântica. Por isso, Won (2005) refere que os metadados capturam a semântica de dados residentes em várias fontes para integração num sistema de informação empresarial.

De acordo com Marco (2013), os metadados foram utilizados como dicionário de dados até aos anos 80. Até aos anos 90, foi utilizado como parte de repositórios para facilitar a utilização de ferramentas CASE. Até 2000, foi utilizado como repositórios específicos para funcionamento de ferramentas de *business intelligence*, no mesmo modelo de funcionamento para as ferramentas CASE. A partir de 2000, tem sido utilizado como forma de tentar capturar e manter informação sobre todos os sistemas das organizações.

Tannenbaum (2001) refere que os riscos de falta de metadados têm impacto ao nível da incorreção da informação, qualidade de dados e perda de produtividade. No domínio específico de *business intelligence* além do problema da existência dos metadados, deve-se considerar igualmente o problema de partilha/troca de metadados quando existem várias ferramentas envolvidas. Por essa razão foi criado o MDC (*Meta Data Coalition*) para definição de padrões, tendo-se criado inicialmente o MDIS (*Metadata Interchange Specification*), sendo que desde 1995 a OMG (*Object Management Group*) tem criado várias especificações de modelação e metadados como é o caso do UML (*Unified Modelling Language*) para modelação, MOF (*Meta Object Facility*) enquanto API para manipulação de metadados, XMI (*XML Metadata Interchange*) para troca de metadados baseado em XML e o CWM (*Common Warehouse Metamodel*) como suporte para representar metadados técnicos e de negócio. Atualmente a OMG suporta uma evolução do CWM designada por IMM (2008).

O conceito de metadados emerge assim com o objetivo inicial de definir de forma abstrata os tipos de dados e relações, nomeadamente no desenho de bases de dados, para garantir uma capacidade de análise prévia ao desenho físico e potenciar opções de implementação que permitam a reutilização e interoperabilidade. Por outro lado, Marco (2013) define igualmente os objetivos dos metadados da seguinte forma:

- Divulgar e gerir os dados da organização;
- Documentar processos;
- Reduzir risco de passagem de conhecimento;
- Analisar e tomar decisões sobre a importância e relevância dos dados.

Pela complexidade de definição no contexto de *business intelligence*, ou com recursos diversos em sistemas de informação (aplicacionais, informacionais, tecnologia), torna-se fundamental o detalhe de normas para definição e reutilização, incluindo interoperabilidade de modelos. De acordo com Marco (2013), as normas devem garantir independência da tecnologia e do fornecedor, mantendo um âmbito específico para a resolução de problemas.

Os metadados devem a sua existência à complexidade e heterogeneidade de arquiteturas aplicacionais e informacionais, o que origina a necessidade de modelação e sistemas para gestão de conhecimento de estruturas de dados, no sentido de permitir gerir a informação e conhecimento sobre as próprias estruturas de dados. Por outro lado, é posicionado como modelo para navegação e estrutura para geração automática de *queries* em ferramentas de *business intelligence*. Daí o interesse em detalhar uma possível classificação de metadados face ao seu objetivo neste domínio específico do *business intelligence* referido por Kimbal e Ross (2013), mas que se pode generalizar para qualquer tipo de aplicação, inclusive para uma organização como um todo:

- **Metadados do sistema:** fontes de dados sobre fontes de dados e processos de extração;
- **Metadados de receção de dados:** dados de auditoria da extração, sobre *schedules*, processamento e registo de dados, sobre tabelas e sobre transformações de campos;
- **Metadados de sistemas de gestão de base de dados:** descrição de bases de dados, tabelas, campos e *scripts*.

Estes tipos são vistos de forma agregada como metadados de negócio (definições de negócio, hierarquias de dados, regras de agregação, métricas), processos (origem, destino, frequência, sequência de processos de carregamento), técnicos (locais físicos, formatos, estruturas) e aplicacionais (como se acede e utiliza, quer por utilizadores quer por aplicações, mas que pode ser visto como parte dos metadados de negócio). Marco (2013) resume os metadados em dois tipos essenciais:

- **Técnicos:** relatórios, frequências de execução, tempo de execução, mapeamentos operacional/analítico, conversões de dados, modelos lógicos/físicos, identificação campos/tabelas/índices/programas e gestão de versões;
- **Negócio:** estruturas de dados numa perspetiva de utilizador, definições de negócio e de atributos mapeados em campos, estatísticas de qualidade de dados, regras de navegação pelos dados, localização/agregação temática dos dados.

Em termos de formas de representação, de entre as normas atuais na indústria, destacam-se as seguintes:

- **MOF (*Meta-Object Facility*):** definido pela OMG (*Object Management Group*), esta norma tinha como objetivo inicial estruturar um metamodelo para o UML (*Unified Modelling Language*), sendo que foi depois tornado como norma pelo ISO/IEC para a definição, gestão e integração de metadados genéricos;
- **OIM (*Open Information Model*):** de acordo com Marco (2013), foi definido como norma em 1996 pela Microsoft e Texas Instruments para interoperabilidade de metadados em *business intelligence*, mas igualmente para processos e regras de negócio. Este modelo foi posteriormente dinamizado pela organização MDC (*Metadata Coalition*), integrado na OMG, mas, entretanto, descontinuado;
- **CWM (*Common Warehouse Metamodel*):** foi definido pela OMG enquanto norma para interoperabilidade tendo como base um metamodelo partilhado, utilizando algumas especificações do OIM, mas focando em sistemas de *business intelligence*. Evoluiu para o IMM (2008);
- **BIDM (*Basic Interoperability Data Model*):** foi definido pelo *Reuse Library Interoperability Group* e pelo *IEEE Software Engineering Standards Committee* tendo gerado a norma IEEE 1420.1 para determinar o conjunto mínimo de ativos de informação em qualquer tipo de *reusable libraries*;
- **DCMI (*Dublin Core Metadados Initiative*):** é definido por um organismo internacional com o mesmo nome, focando na especificação de elementos de metadados para descrição de recursos para permitir a criação de sistemas inteligentes na descoberta/exploração desses dados, tendo originado a norma ISO 15836-2003;
- **ESM (*Enterprise System Metamodel*):** segundo Marco (2013), foi definido por empresas de consultoria para detalhar todas as estruturas de dados de uma organização e estabelecer a sua relação com metadados;
- **ITPMM (*IT Portfolio Management Metamodel*):** de acordo com Marco (2013), foi definido por empresas de consultoria tendo por base ideia de que os ativos de sistemas de informação como projetos, software, hardware, pessoas e serviços, são fundamentais numa organização e têm estruturas próprias descritivas e de relação entre elas, que podem ser generalizadas.

Um dos modelos mais recentes é o IMM (2008), gerido pela OMG que tem outras especificações no domínio da interoperabilidade técnica como o CORBA, MDA (*Model Driven Architecture*) e BPMN (*Business Process Management Notation*). O modelo do IMM recupera definições consideradas igualmente por Marco (2013) e Tannenbaum (2001) para apresentar os metadados decompostos em conceitos técnicos e de negócio, além de que indica que existem metamodelos que são metadados sobre os próprios metadados. Por outro lado, a especificação IMM recorre ao referencial de Zachman para arquiteturas pois torna-se necessário enquadrar o IMM num modelo alargado de informação.

Tendo por base a definição de tipos de metadados e referencial de Zachman, o IMM ilustra a diferença entre dados e metadados tendo por base um sistema de exemplo de uma ATM (*Automatic Teller Machine*) como apresentado na Tabela 3.2, onde se ilustram vários exemplos de dados instanciados e como são considerados como metadados.

This Book (Meta- metadata)	Elements of metadata (metadata model)	Objects: "Entity Class", "Attribute"	Objects: "Entity Class" "Attribute" "Role"	Objects: "Table" "Column"	Object: "Program module", "Language"
Data Management (Metadata)	Data about a database (a data model)	Entity class: "Customer" Attributes: "Name" "Birthdate"	Entity classes: "Branch", "Employee" Attributes: "Employee.Address" "Employee.Name" Role: "Each Branch must be managed by exactly one Employee"	Table: "CHECKING_ ACCOUNT" Columns: "Account_number" "Monthly_charge"	Program module: ATM Controller Language: Java
IT Operations (Instance Data)	Data about real world things (a database)	Customer Name: "Julia Roberts"; Customer Birthdate: "10/28/87"	Branch Address: "111 Wall Street" Branch Manager: "Sam Sneed"	CHECKING_ ACCOUNT. Account_number: = "09743569" CHECKING_ ACCOUNT. Monthly_charge: "\$4.50"	ATM Controller: Java code
	Real world things	Julia Roberts 	Wall Street branch 	Checking account #09743569 	ATM Withdrawal 

Tabela 3.2: Relação entre dados, bases dados, modelos de dados e metadados (IMM, 2008)

Tendo por base este enquadramento, o IMM posiciona-se no centro de modelação de negócio, bases de dados, aplicações e outros modelos incluindo ontologias e capacidade de *data lineage* (designado por *traceability metamodel*), como apresentado na Figura 3.16.

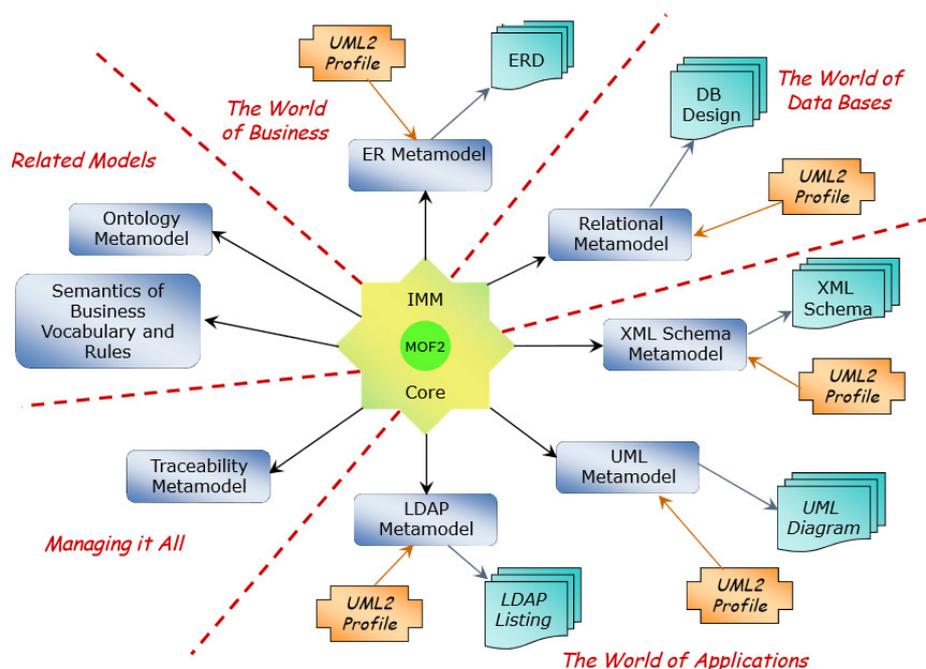


Figura 3.16: Modelo IMM vs outros modelos (IMM, 2008)

Para garantir a capacidade de integração de vários metamodelos, o IMM define uma estrutura de componentes apresentado na Figura 3.17, assente na captura de metamodelos (*model management*) e numa estrutura base (*IMM Core*), a partir do qual se enquadra os metadados de negócio (*business modeling*) e técnicos (*technology modeling*). Cada componente é depois objeto de especificação detalhada em modelos UML (IMM, 2008).

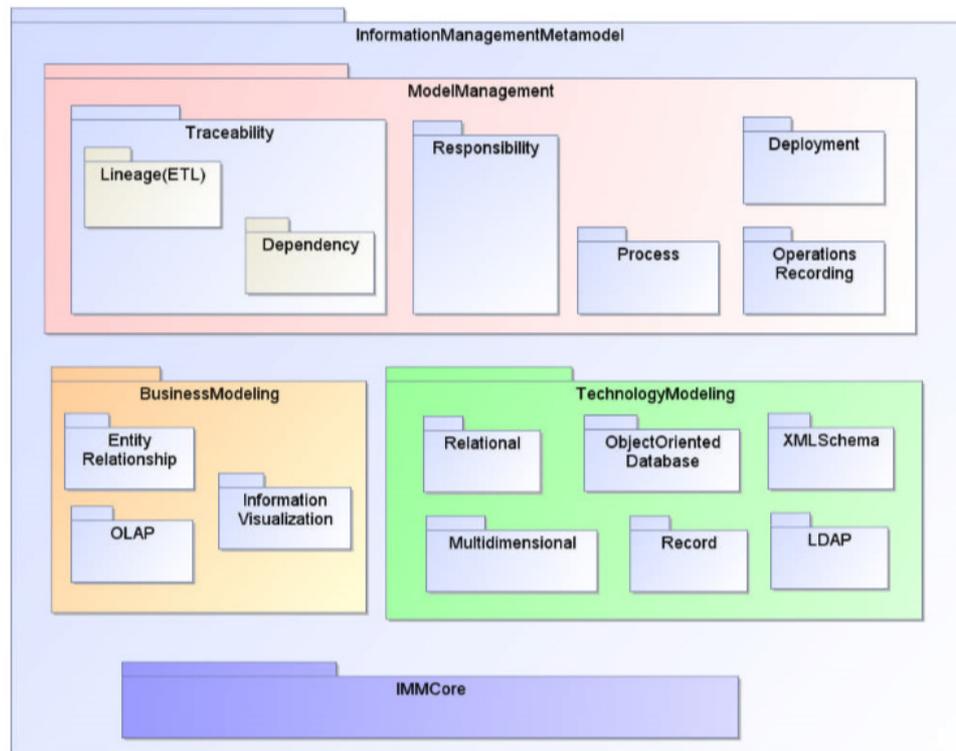


Figura 3.17: Âmbito da especificação IMM (IMM, 2008)

De acordo com Tannenbaum (2001), o repositório de metadados corresponde a uma ferramenta que define e gere as definições de dados de uma organização, bem com as suas relações, tendo como requisitos mínimos uma base de dados, metamodelo e *software* para gestão. Como tal, existem várias ferramentas enquanto repositório para metadados que implementam vários dos metamodelos enquanto normas, como as seguintes:

- **RIB (*Repository in a box*):** é uma ferramenta *open source* desenvolvido pelo *Innovative Computing Laboratory* da Universidade do Tennessee, enquanto implementação de metadados como informação de objetos reutilizáveis, suportando o modelo BIDM e utilizando tecnologia MySQL e Java. Atualmente está descontinuada;
- **DSPACE:** é uma ferramenta *open source* para capturar, registar, indexar e distribuir material digital de pesquisa, desenvolvido pelo MIT em parceria com a HP. O material de pesquisa é considerado ao nível de artigos, relatórios técnicos, artigos técnicos, teses, em formato multimédia. Utiliza o modelo DCMI com base de dados PostGres ou Oracle, e plataforma em Java;
- **SEIS *Metadada Repository*:** foi desenvolvido pelo *Space Environment Information* que pertence à Agência Espacial Europeia, tendo como objetivo a otimização das operações de voo, utilizando tecnologia Microsoft “.Net” e Microsoft SQLServer para implementação do SEIS (*Metadata Repository*) enquanto metamodelo próprio.

No entanto, as ferramentas de *business intelligence* mais comerciais segundo a Gartner (2015), têm modelos próprios de gestão de metadados integrados nas suas soluções de *data integration*, como é o caso da IBM, SAS, Oracle e Informatica, com capacidade de integração de metadados externos a partir de conectores proprietários. Estas soluções são compostas por repositórios proprietários sobre o qual se disponibilizam ferramentas para exploração via *data lineage*, *data discovery* e integração de metadados por *workflow* no caso específico de metadados de negócio onde é necessário enriquecer os metadados inicialmente gerados pelas ferramentas com descrições mais adequadas para os utilizadores.

No caso específico da Microsoft, existe um projeto *open source* Microsoft SQLServer Metadata Toolkit [SQLToolkit] (2016) que permite integrar metadados a partir dos vários metadados gerados pelas soluções de *business intelligence* (SQLServer, Integration Services, Analysis Services). Esta aproximação desta solução, mostra a necessidade de integração e consolidação a partir de vários objetos de metadados gerados pela sua tecnologia ao nível de bases de dados, modelos analíticos, processos de ETL e mesmo *reporting*, tendo por base uma relação entre os conceitos para garantir uma visão integrada de metadados. Utiliza conectores próprios às várias ferramentas Microsoft, interpretando as representações de metadados em cada ferramenta, para consolidar estes metadados. Com base na visão de metadados consolidada, permite a visualização sobre a forma de grafo ou descritiva de cada conceito de metadados e das suas relações, tal como apresentado.

3.3.3.3 Ontologias

As ontologias permitem criar um modelo semântico de metadados baseado em tripletos de objeto-relação-objeto, sobre o qual é possível utilizar motores de inferência num domínio.

Segundo Gruber (1993) as ontologias são especificações formais (para ser lida por máquinas) e explícitas (conceitos e restrições explicitamente definidos) de uma conceptualização partilhada (ideia do mundo para um grupo de pessoas). Por outro lado, Jasper e Uschold (1999) definem ontologia como um conceito que pode ter várias formas, mas que inclui um vocabulário de termos e especificação do seu significado. Segundo Heflin (2004), o conceito de ontologias tem sido utilizado para descrever vários tipos de artefactos, incluindo taxonomias, esquemas de metadados do DCMI e mesmo teorias, o que comprova a sua utilidade com diversas finalidades, mas que se refletem em metodologias, ferramentas e especificações.

As ontologias permitem capacidade expressiva sobre tipo de objetos ou conceitos num determinado domínio, atributos desses conceitos e relações onde esses conceitos estão envolvidos, permitindo instanciar a ontologia criando uma base de conhecimento, como referido por Noy e McGuinness (2001). Os seus componentes são os seguintes, de acordo com Gomez-Perez et al. (2006):

- **Classes:** correspondem a conceitos, sendo utilizados em taxonomias, representando tarefas, ações, estratégias, processos de raciocínio, entre outros conceitos. Uma classe corresponde a uma estrutura que representa um conjunto de objetos ou instâncias que assumem essa estrutura;
- **Relações:** representam tipos de interação entre as classes de um domínio, determinando uma classificação dessas relações enquanto “subclasse de”, “conectada a”. Pode ser uma sub-classe ou uma relação entre classes com um predicado;

- **Funções:** tipo especial de relação onde o último elemento da relação pode ser visto como um elemento determinístico face aos elementos precedentes;
- **Axioma:** definem o significado e restrições, que permitem modelar expressões sempre verdadeiras;
- **Instâncias:** representam elementos específicos (os dados que são instanciados).

Segundo Noy e McGuiness (2001), as ontologias têm um papel importante na reutilização e partilha de conhecimento pelas seguintes razões:

- Partilha de entendimento sobre a estrutura da informação entre pessoas e *software*;
- Permitir reutilização de conhecimento entre domínios, racionalizando o esforço de desenvolvimento de ontologias ao permitir reutilizar parte ou a totalidade entre si;
- Tornar explícito as suposições, clarificando entendimento de conceitos;
- Permitir uma separação lógica entre conhecimento de domínio e conhecimento operacional, permitindo implementar visões operacionais a partir de conhecimento do domínio, mantendo uma base comum de entendimento dos conceitos;
- Permite uma análise de conhecimento de domínio com base na análise das especificações dos termos constantes nas ontologias.

Uma ontologia representa um domínio de conhecimento específico do mundo sobre a forma de conceitos, propriedades e relações tendo por base três tipos de informação, de acordo com Bézvin (1998):

- **Terminológica:** conceitos e relações utilizadas para a definição da ontologia (e.g. Cliente, Conta);
- **Assertiva:** axiomas da ontologia enquanto assertivas que se aplicam aos conceitos e relações (e.g. um cliente pode ser particular ou empresa; uma conta é um contrato);
- **Pragmática:** forma de apresentação dos conceitos relações recorrendo a ferramentas como editores ou inferência.

Bieman (2005) cita Sowa que classifica as ontologias em três tipos de acordo com Figura 3.18:

- **Formal ontology:** concetualização de categorias que se podem distinguir por axiomas e definições baseadas em lógica para suporte a inferência;
- **Prototype-based ontology:** categorias correspondem a protótipos de instâncias do mesmo tipo;
- **Terminological ontology:** é o caso do *WordNet* que corresponde a conceitos relacionados com modelos de subtipo-supertipo.

Formal ontology

Axioms:

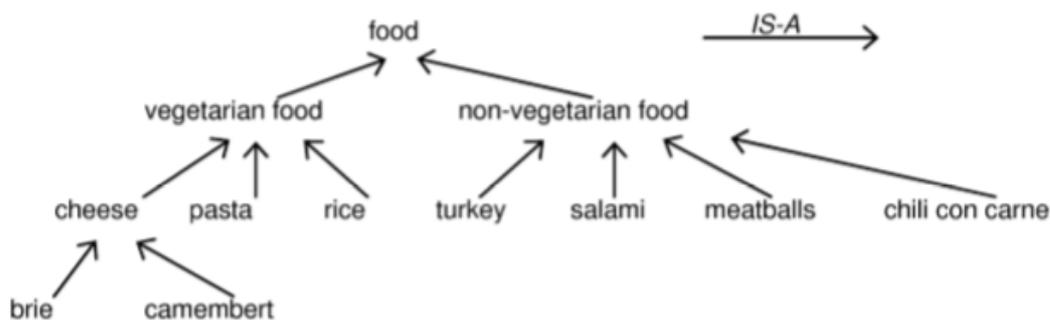
food(brie), food(camembert), food(turkey), food(meatballs), food(chili con carne), meat(turkey), meat(minced meat), part_of(minced meat, chili con carne), part_of(minced meat, meatballs)

veg_food(x) = { x | food(x) ∧ (¬part_of(y,x) ∧ meat(y)) ∧ ¬meat(x) }

non_veg_food(x) = { x | food(x) ∧ ((part_of(y,x) ∧ meat(y)) ∨ meat(x)) }

Possible to derive: "turkey" and "chili con carne" are non-vegetarian foods

Terminological ontology



Prototype-based ontology



Figura 3.18: Tipologia de ontologias segundo John Sowa (Bieman, 2005)

Segundo Morais e Ambrosio (2007), quanto ao grau de formalismo, as ontologias podem ser categorizadas em altamente informais (quando expressas em linguagem natural; semi-informais, quando expressas em linguagem natural de forma restrita e estruturada; semi-formais, expressas em linguagem artificial definida formalmente) ou rigorosamente formais (os termos são definidos com semântica formal, teoremas e provas). O mesmo autor categoriza as ontologias nos seguintes tipos de acordo com a sua função ou objetivo:

- **Ontologias genéricas:** descrevem conceitos mais amplos, como elementos da natureza, espaço, tempo, coisas, estados, eventos, processos ou ações, independente de um problema específico ou domínio particular;
- **Ontologias de domínio:** descrevem conceitos e vocabulários relacionados a domínios particulares, tais como medicina ou computação, sendo o tipo de Ontologia mais comum, geralmente construída para representar um domínio particular;

- **Ontologias de tarefas:** descrevem tarefas ou atividades genéricas, que podem contribuir na resolução de problemas, independente do domínio que ocorrem (e.g. processos de vendas ou diagnóstico). A sua principal motivação é facilitar a integração dos conhecimentos de tarefa e domínio em uma abordagem mais uniforme e consistente;
- **Ontologias de aplicação:** descrevem conceitos que dependem tanto de um domínio particular quanto de uma tarefa específica. Devem ser especializações dos termos das ontologias de domínio e de tarefas correspondentes;
- **Ontologias de representação:** explicam as conceitualizações que fundamentam os formalismos de representação de conhecimento, procurando tornar claros os compromissos ontológicos embutidos nestes formalismos, como é o caso exemplo da ontologia de *frames* utilizada em Ontolíngua.

Para tipificar as ontologias, existem várias formas, sendo de ressaltar a definição de Gruber (1993) que publicou um referencial específico no domínio da ciência da computação e informação, agrupado por dimensões:

- **Dimensão semântica:** relacionada com a especificação do vocabulário. Pode ser decomposta pelos seguintes níveis:
 - **Nível de estrutura:** também visto como nível de formalidade ou nível de dados estruturados, sendo como tal uma ontologia “altamente estruturada” (bem especificada tal como no caso de abstrações matemáticas), “pouco estruturada” (generalista nos conceitos como é o caso de documentos) e “semiestruturada” (misto de definições formais e informais);
 - **Expressividade da linguagem:** relacionada com o nível de estrutura, a expressividade da linguagem em ontologias altamente estruturadas exige expressividade e formalismo, enquanto que baixo nível de estrutura exige no limite uma linguagem que permita listagens de condições e definições com baixa expressividade formal;
 - **Granularidade representacional:** nível de detalhe dos conceitos.
- **Dimensão pragmática:** relacionada com a finalidade e contexto da ontologia. Pode ser decomposta pelos seguintes níveis:
 - **Intenção de uso:** qual o grupo de utilização da ontologia, considerando representação de vocabulários em língua natural, partilha de bases de conhecimento, modelo de comunicação entre agentes de *software*;
 - **Papel da automação lógica:** possibilidade de utilização de automação lógica para inferência a partir dos conceitos e relações;
 - **Descritiva versus prescritiva:** ontologias com notações mais descritivas e livres, ou mais prescritivas em termos de rigidez na caracterização e declaração de objetos/classes;
 - **Metodologia:** as ontologias podem ser construídas com metodologias *bottom-up* a partir dos conceitos base e depois agregando e relacionado ao nível de subclasses e classes do domínio, ou *top-down* onde se inicia pela caracterização do domínio e depois se efetua a decomposição até às classes mais básicas.

Para a implementação das ontologias, existem assim várias linguagens, algumas mais tradicionais e outras focadas para explorar as características da *World Wide Web* e normalmente baseadas em XML. De acordo com Gomez-Perez et al. (2006) a classificação pode ser a seguinte:

- **Traditional ontology:** caso de ontologia baseada em *first-order predicate* KIF (Knowledge Interchange Format) e reutilizada em várias outras linguagens como LOOM e CYCL; FLogic (*frame logic*) baseada em frames e *first-order predicate*; Cycl baseada em *first-order Predicate*; Loom baseada em *description logic*;
- **Web Standard and Recommendation:** linguagens criadas no contexto da web semântica como o RDF, complementada com RDFS para permitir definir recursos como conceitos e relações entre propriedades e recursos;
- **Web-Based Ontology:** linguagens como a XOL (*XML-Based Ontology Exchange Language*), SHOE (*Simple HTML Ontology Extension*), OML (*Ontology Markup Language*), OIL (*Ontology Interchange Language*), DAML-OIL (*DARPA Agent Markup Language – Ontology Interchange*) e OWL.

Existem ainda linguagens que assumem uma importância no domínio da web semântica e na sua aplicabilidade atual, centrada no OWL, mas que deriva de uma evolução do mesmo tipo de linguagens, como é o caso das seguintes:

- **SHOE:** corresponde a extensões de HTML para se obter conteúdo semântico em páginas Web;
- **OIL:** combina formalismo de *frames* com semântica formal e inferência lógica descritiva para classificação e taxonomia de conceitos;
- **DAML+OIL:** evolução da linguagem OIL combinada com DAML e SHOE determinando a normalização final ao nível do OWL;
- **RDF:** foi desenhado como norma pela W3C para descrição de recursos ou objetos através de uma categorização de recursos ou objetos que têm várias propriedades e estas têm valores, como numa lógica de predicados utilizada em linguagem natural;
- **OWL:** foi criado pela W3C como uma extensão do RDF, acrescentado mais termos de vocabulário para descrever as propriedades, como é o caso de cardinalidade, controlo, tipos de dados mais complexos, classes de dados enumeradas, entre outros conceitos.

No caso específico de regras existem ainda várias linguagens específicas como as seguintes:

- **RuleML:** *rule markup language* para definir mecanismos de *forward* e *backward* sobre representações em XML, tendo servido de base para o SWRL;
- **SWRL:** combinação de RuleML com OWL Lite, utilizando conceitos de *Horn Logic* com OWL. Esta linguagem tem hoje utilização em ferramentas como Protégé e motores de inferência como Pellet.

De acordo com Pereira (2012), existem várias metodologias para criar ontologias, como é o caso da DOGMA-MESS, On-To-Knowledge ou CommonKADS1, destacando-se no entanto a metodologia Methontology que é recomendada pela FIPA (*Foundation for Intelligent Physical Agents*) e que estabelece as seguintes fases:

- Especificar os requisitos;
- Conceptualizar o domínio do conhecimento;
- Formalizar o modelo conceptual numa linguagem formal;
- Implementar um modelo formal;
- Fazer a manutenção das ontologias implementadas.

Como atividades de suporte a esta metodologia, são referidas: aquisição do conhecimento, integração, avaliação, documentação e gestão da evolução. Pereira (2012) refere ainda a metodologia DILIGENT (*Distributed, Loosely-controlled and Evolving Engineering of Ontologies*), que foi usada numa organização virtual sobre o projeto europeu SWAD (*Semantic Web Advanced Development*), sendo a primeira a colocar ênfase, não no desenvolvimento inicial da ontologia, mas sim no utilizador e na sua utilização da ontologia, bem como, nas mudanças introduzidas por este. Esta metodologia considera cinco fases:

- **Construção:** processo começa tendo os peritos de domínio, utilizadores, engenheiros do conhecimento e engenheiros de ontologias construído uma ontologia inicial. Em contraste às outras metodologias de engenharia de ontologias, não é requerido a completude da ontologia inicial. O grupo envolvido na construção da ontologia inicial deve ser relativamente pequeno, de forma a encontrar uma primeira versão consensual da ontologia partilhada, numa forma fácil;
- **Adaptação local:** uma vez que a ontologia central esteja disponível, os utilizadores trabalham com ela e adaptam-na localmente às suas necessidades. No seu ambiente local, são também livres para alterar a ontologia central reutilizada. No entanto, não lhes é permitido diretamente mudar a ontologia central. A central de controlo recolhe pedidos de mudança para a ontologia partilhada e guarda as adaptações locais (continuamente ou em pontos de controlo);
- **Análise:** a central de controlo analisa as ontologias locais e os pedidos para mudanças para tentar identificar semelhanças nas ontologias dos utilizadores. Dado que nem todas as mudanças introduzidas ou pedidas pelos utilizadores serão introduzidas na Ontologia central partilhada, a atividade crucial da central de controlo está em decidir que mudanças são introduzidas na próxima versão da ontologia partilhada;
- **Revisão:** a central de controlo deve rever regularmente a ontologia partilhada, de modo a que as ontologias locais não diverjam demasiado da ontologia partilhada. O objetivo da revisão é realinhar a ontologia com as necessidades dos utilizadores e assim, ganhar elevada aceitação e menor diferenciação local. Os utilizadores são envolvidos na avaliação da ontologia sob o ponto de vista da usabilidade. Os peritos de domínio são responsáveis por avaliá-la do ponto de vista do domínio. Os engenheiros do conhecimento são, principalmente, responsáveis pela avaliação a nível de domínio e técnico (e.g. eficiência, conformidade com as normas). Os engenheiros da ontologia também têm responsabilidade sob o ponto de vista técnico. Outra tarefa da central de

controlo é assegurar a compatibilidade com as versões anteriores. Os engenheiros da ontologia são responsáveis por atualizá-la, baseados nas decisões da central de controlo. A revisão da ontologia partilhada compreende a sua evolução;

- **Atualização local:** uma vez que a nova versão da ontologia partilhada seja disponibilizada, os utilizadores podem atualizar as suas próprias ontologias.

Morais e Ambrosio (2007) referem igualmente as metodologias Methontology e On-to-knowledge, além da metodologia Enterprise, para apresentar uma metodologia consolidada, baseada nas seguintes fases e com documentação em todas as fases:

- **Propósito e especificação de requisitos:** identificar a competência da ontologia, ou seja, seus usos e propósitos, através da delimitação do que é relevante para a ontologia e o que não é. Além disso, nesta fase, podem ser identificados potenciais utilizadores da ontologia e o contexto que motivou sua construção. Após definir a competência, finalmente os requisitos da ontologia devem ser especificados. Essa especificação envolve a descrição do propósito e dos usos da ontologia, isto é, a definição das questões que a ontologia deve ser capaz de responder (questões de competência). Estas questões, além de justificar a existência da ontologia, servem para futura avaliação da mesma;
- **Captura da ontologia:** considerada a fase mais importante, seu objetivo é capturar o conjunto de elementos de um domínio que podem ser representados em uma ontologia, com base nas questões de competência. A captura envolve a identificação e especificação de conceitos (classes), seus relacionamentos e todos os demais elementos necessários para a representação da ontologia, tais como propriedades, axiomas, instâncias. Alguns conceitos que não podem ser obtidos a partir de outros conceitos da ontologia (conceitos primitivos) devem ser especificados utilizando-se linguagem natural, com o mínimo de ambiguidade e inconsistência possível. Já os conceitos não primitivos devem ser especificados com referência clara aos conceitos primitivos. A utilização de taxonomias (organização em categoria e subcategorias) também deve ser utilizada na representação dos conceitos, pois auxilia sua organização, facilitando seu entendimento;
- **Formalização da ontologia:** especificação da ontologia em uma linguagem. Na prática, uma ontologia pode ser representada através de qualquer linguagem não formal (natural), embora a representação formal seja considerada a mais apropriada na maioria dos casos. Uma linguagem formal é uma linguagem fundamentada em um modelo matemático. Geralmente envolve pressuposições implícitas que entram despercebidas no processo de dedução, tem símbolos não ambíguos e formulações exatas. Como tal, a clareza e a correção de uma dedução podem ser testadas com maior facilidade e precisão. O que determina a escolha da linguagem a ser utilizada na formalização da ontologia depende do propósito da ontologia a ser desenvolvida;
- **Integração com ontologias existentes:** durante os processos de captura e/ou formalização, pode surgir a necessidade de integrar a ontologia em questão com outras já existentes, visando aproveitar conceitualizações previamente estabelecidas;
- **Avaliação:** verificar se a mesma satisfaz os requisitos definidos em sua construção, com base em critérios (clareza, coerência, extensibilidade e compromissos ontológicos mínimos) para guiar tanto o desenvolvimento, quanto a avaliação da qualidade das ontologias construídas.

Para Morais e Ambrosio (2007) no contexto as metodologias que ferem, os termos capturados devem ser descritos por um dicionário de termos, considerando dois princípios importantes a seguir descritos, sugerindo a utilização de hipertexto para a documentação, tendo por base estes princípios:

- **Princípio do vocabulário mínimo:** diz respeito ao vocabulário utilizado na definição dos termos da ontologia. Deve ser o menor possível e não deve apresentar ambiguidades. Qualquer termo que não tenha um significado claro e não ambíguo, deve ser formalmente definido no dicionário;
- **Princípio da autorreferência:** indica que a definição de um termo no dicionário deve, sempre que possível, ser feita utilizando outros termos do mesmo.

Para Noy e McGuinness (2001) existem sete passos fundamentais para a construção de ontologias:

- Determinar o domínio e âmbito;
- Avaliar a reutilização de ontologias existentes;
- Enumerar os termos importantes;
- Definir as classes a sua hierarquia;
- Definir as propriedades das classes;
- Definir as restrições das propriedades;
- Criar as instâncias das classes.

Para facilitar a construção e utilização das ontologias, têm sido desenvolvidas várias ferramentas que permitem reduzir o nível de conhecimento exigido sobre as especificações de linguagens, utilizando-se interfaces mais amigáveis com o utilizador. De notar, no entanto, que a maioria das ferramentas utiliza como base a especificação XML pois esta proporciona uma sintaxe para documentos estruturados, mas impõe a não existência de restrições semânticas no significado desses documentos, pois é uma sintaxe genérica. É o caso das seguintes ferramentas:

- **Protégé:** criado pela Stanford University em regime de código aberto, funciona como um editor de ontologias com *plugins* de interface com FLogic, JESS, OIL, OWL e RDF;
- **WebODE:** desenvolvida pelo Laboratório de Inteligência Artificial da Universidade Técnica de Madrid numa arquitetura Web que permite interfaces com RDF, OIL, DAML, FLOGIC e JESS;
- **OntoEDIT:** desenvolvida pelo AIFB da Universidade de Karlsruhe, tem capacidade de interface via *plugins* com FLOGIC, RDF e DAML-OIL;
- **Altova Semantic Works:** ferramenta comercial da empresa Altova, é um editor RDF e OWL para web semântica;
- **NEON:** criado pelo projeto Europeu Neon IST-2005-027595 sobre Eclipse para criação de Ontologias em rede criadas de forma colaborativa;
- **OWLEditor:** propriedade da empresa Model Futures que esteve envolvida igualmente no desenvolvimento do referencial MODAF.

3.3.3.4 Web semântica

A web semântica (Berners-Lee et al., 2006) é uma forma de estender o potencial da Web, passando de uma visão de documentos partilhados, para dados partilhados. Corresponde a uma extensão das linguagens de suporte à Web tendo como objetivo desenvolver padrões, arquiteturas de metadados e linguagens de ontologias para permitir processamento e integração de informação de forma automática por computadores. Para o efeito, como referido em Berners-Lee et al. (2006), considera-se uma arquitetura integrando várias tecnologias em camadas apresentada na Figura 3.19 que permite descrever a informação através de estrutura e semântica a partir de identificadores únicos de informação de base.

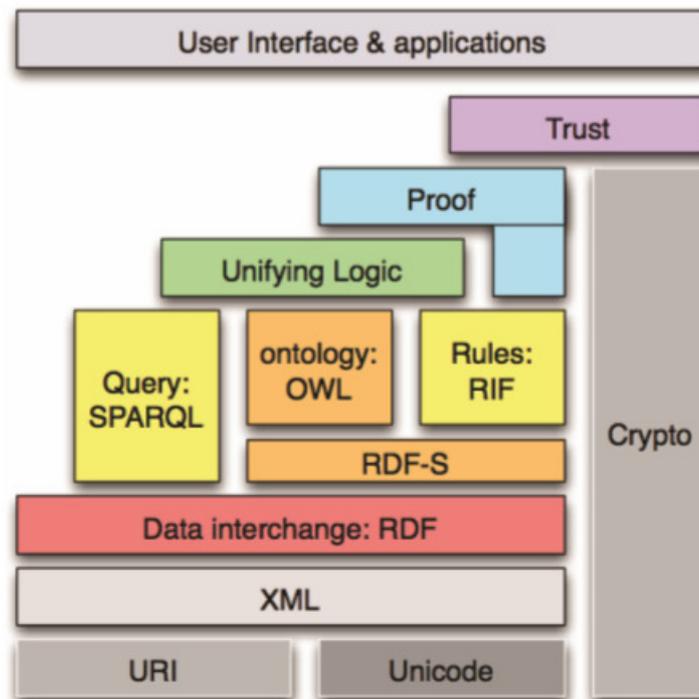


Figura 3.19: Pilha tecnológica web semântica (Berners-Lee et al., 2006)

Esta arquitetura, descrita em Berners-Lee et al. (2006) pode ser descrita da seguinte forma, de acordo com a nossa interpretação:

- **Camada de identificação da informação:** identificadores URI (*Uniform Resource Identifier*) e caracteres padronizados UNICODE que permitem criar chaves únicas para conceitos como classes e propriedades;
- **Camada de estrutura da informação:** utiliza a sintaxe base XML, incluindo tipos de dados e espaço de nomes (*name spaces*), para permitir estruturar os conceitos;
- **Camada de semântica da informação:** suportado em linguagem RDF para relacionar conceitos como grafos simples, RDFS como extensão de RDF com primitivas baseadas em *frame*, ou OWL para permitir capacidade de descrição lógica (classes disjuntas, propriedades de relação e propriedades de dados);
- **Camada de exploração de informação:** utiliza linguagens como SPARQL para interrogar modelos, mas podemos igualmente considerar a inferência sobre os modelos de ontologia. Permite igualmente criar as interfaces para desenvolvimento de várias aplicações assentes na estrutura e semântica da informação base, vista como ontologia.

Considerando a pilha tecnológica apresentada na Figura 3.19, destacamos para a nossa investigação a forma de representação de ontologia via OWL e a linguagem de interrogação de ontologia via SPARQL a seguir detalhado.

A linguagem OWL da World Wide Web Consortium [OWL] (2013) permite descrever classes e relações, sendo uma revisão do DAML+OIL e evolução sobre o RDF. No caso do OWL divide-se em três alternativas também designado por módulos:

- **OWL Lite:** utiliza a mesma definição de classes, propriedades e hierarquias presente no RDF, além de implementar semântica para definir propriedades de transitividade, simetria, relação inversa e igualdade entre classes e relações. Suporta utilizadores que necessitam principalmente de uma classificação hierárquica e de características de restrições simples;
- **OWL DL (*Description Logic*):** permite definir classes por união, interseção e complemento, além de enumeração de instâncias e disjunções. Suporta igualmente a definição de relações de instâncias e literais RDF. Suporta os utilizadores que querem expressividade máxima, sem perderem a totalidade de capacidade de decisão computacional dos sistemas de raciocínio;
- **OWL Full:** suporta cardinalidade máxima e mínima de propriedade, além de permitir criação de metaclasses. Suporta os utilizadores que querem expressividade máxima e liberdade sintática de RDF, sem garantias computacionais.

Tal como o XML, o OWL utiliza *namespaces*, onde se define um conjunto de estruturas tal como apresentado na Figura 3.20.

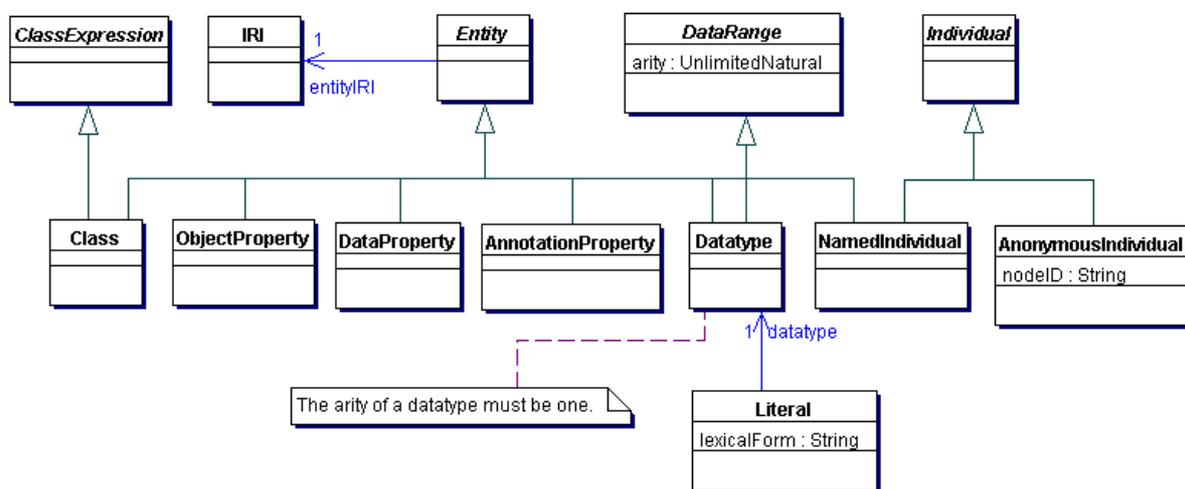


Figura 3.20: Estrutura de uma ontologia OWL (OWL, 2013)

De acordo com este modelo, destaca-se o conceito de IRI (*Internationalized Resource Identifiers*) que permite identificar um recurso de forma única dentro do *NameSpace* (e.g. <http://www.empresa.org/ontologia#produto>, no caso de se querer definir uma classe produto, no *namespace* <http://www.empresa.org/ontologia>). Tendo por base este conceito chave, criam-se modelos de ontologia, a partir do seguinte, sendo que a base de regras é indicada como *statement* numa estrutura de sujeito-predicado-objeto:

- **Class:** corresponde normalmente ao sujeito ou entidade. Numa frase pode ser o sujeito, advérbio ou um substantivo, que representa uma entidade real;

- **Data property:** corresponde a uma propriedade de uma entidade, que está associado normalmente a um valor, designado como literal. Numa frase, corresponde ao adjetivo que descreve o sujeito;
- **Object property:** relações ou predicados, que caracterizam a ligação entre duas entidades. Pode relacionar duas *class*, ou uma *class* com uma *Data Property*. Numa frase, é o verbo pois a ação estabelece uma relação.

A linguagem SPARQL da *World Wide Web Consortium* [SPARQL] (2014) foi criada inicialmente para manipular dados representados sobre a forma de RDF/OWL como forma de pesquisar ontologias. Tal como SQL para bases de dados relacionais, o SPARQL permite selecionar atributos, efetuar junções, ordenações, restrições e aplicar funções sobre a ontologia, com base na seguinte estrutura de linguagem, sendo que efetua a correspondência entre sujeito-predicado-objeto, com base nos *statement* equivalentes na ontologia.

```
PREFIX <Ontologia>:<URI>  
FORM ?<atributo>, FUNCTION(<atributo>)  
WHERE { ?<sujeito> ?<predicado> ? <objeto>}  
GROUP BY ?<atributo> HAVING ?<condição>  
ORDER BY ?<atributo>
```

O “<atributo>”, pode ser um sujeito, predicado ou objeto. O construtor FORM pode ser SELECT (retorna a lista de sujeito-predicado-objeto que corresponde às restrições indicadas em WHERE), CONSTRUCT (retorna um grafo RDF), ASK (retorna um valor verdadeiro ou falso de acordo com a restrição indicada em WHERE) e DESCRIBE (retorna um grafo RDF com os metadados dos recursos encontrados).

3.3.3.5 Bases de dados

Os modelos de dados correspondem a coleção de conceitos que podem ser utilizados para descrever dados, suas relações, sua semântica e restrições de acordo com Elmasri e Navathe, citado por Alalwan (2011), referindo que por sua vez uma base de dados pode ser vista como uma forma de representação de aspetos específicos do mundo (universo de discurso), sendo uma coleção coerente de factos com algum significado.

Os sistemas de gestão de bases de dados são componentes de *software* que implementam modelos de dados, de acordo com determinadas categorias. A categoria mais utilizada são os modelos de dados relacionais sendo que mesmo ao nível de dados disponíveis via internet, perto de 77% dos mesmos estão armazenados em bases de dados relacionais de acordo com Silberschatz citado por Alalwan (2011). No entanto, existem outras categorias de bases de dados, nomeadamente:

- **Modelo em rede e hierárquico:** são modelos de bases de dados que já não são muito utilizadas. Além de modelos em grafos apresentado por Angles e Gutierrez (2008), destaca-se o modelo hierárquico *que* define uma estrutura de relações entre registos utilizando ponteiros para navegar entre os dados, sendo que os ponteiros implementam

o conceito de relações com as restrições de não poder existir mais do que uma relação entre os registos, tal como apresentado por Tsichritzis e Lochovsky (1976);

- **Modelo por objetos:** tendo por base o conceito de linguagens orientadas a objetos centrada no conceito de classes e seus atributos, surgiram desde os anos 70 várias implementações de bases de dados deste tipo focadas na criação de persistência de objetos instanciados pelas linguagens e permitindo a sua utilização com linguagens como Smaltalk, C++ e Java, mas igualmente com linguagens como LISP, JADE, entre outras;
- **Modelo semântico:** bases de dados que permitem reutilizar o conceito de ontologia que captura classes e relações entre classes, além de atributos de classes e regras de negócio (semântica) (Hammer e McLeod, 1981; Alawan, 2011). Em particular, tirando partido de linguagens de representação de ontologias no domínio da web semântica como o RDF e OWL e linguagens de exploração como SPARQL, criaram-se alternativas para a implementação deste tipo de bases de dados como é o caso de GraphDB (2016) ou adaptação de bases de dados relacionais para implementar este conceito como é o caso de Microsoft SQLServer Semantic DB [SQLServerSemanticDB] (2014);
- **Modelo informacional:** bases de dados com uma organização específica para tirar partido de motores OLAP (*Online Analytical Processing*) no domínio de *business intelligence*, implementando modelos de dados em *star* ou *snowflake* (Kimbal e Ross, 2013).

Para a investigação iremos considerar as bases de dados relacionais aqui descritas. No caso de base de dados semântica, pelo facto de serem implementadas via ontologias em particular modelos de OWL e RDF com suporte de linguagens SPARQL, não são relevantes pois o objetivo da investigação a este nível é adequação de ontologias com estas linguagens, mesmo que não estejam suportadas em bases de dados em termos de persistência. No caso de bases de dados informacionais, estas estão descritas especificamente no sub-capítulo 3.4 relativo a sistemas de informação, na seção específica sobre sistemas de *business intelligence* onde estes tipos de bases de dados são utilizados.

Alalwan (2011) cita Date, e Elmasri e Navathe, para definir que uma base de dados é construída sobre um modelo de dados, que corresponde a conceitos que permitem descrever dados, relações semântica dos dados e restrições, sendo que esses dados estão persistentes, isto é, são guardados em memória física, não sendo como tal, dados de entrada e saída entre sistemas.

Os sistemas de gestão de bases de dados relacionais, concetualizados por Codd (2002), implementam conceitos de base de dados, esquema, tabelas, campos e relações entre tabelas, além de conceitos de *views*, *stored procedures* e índices, tendo por base os dados organizados em tabelas composta por linhas (registos) e colunas (campos). Alalwan (2011) cita Elmasri e Navathe, e Silberschatz et al., para definir este modelo considerando que existem as componentes de entidades, atributos e relações. Vilannueva-Rosales (2011) cita igualmente Elmasri e Navathe, mas considerando que existem entidades definidas como tabelas, colunas enquanto atributos e as restrições enquanto relações. Ambos os autores, consideram conceitos de normalização de dados enquanto primeira, segunda e terceira forma normal, enquanto forma de modelação de uma base de dados, além de referirem a linguagem SQL para criar e explorar este tipo de bases de dados. Este modelo é representado por Astrova (2009) na Figura 3.21.

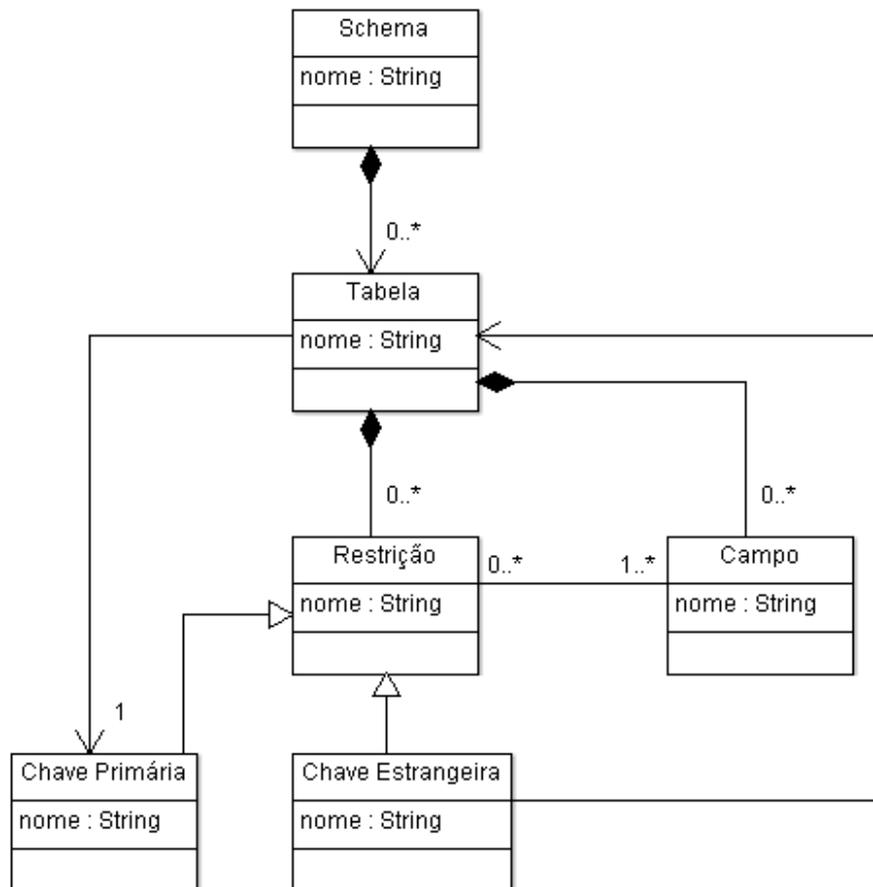


Figura 3.21: Modelo Relacional, adaptado de Astrova (2009)

No caso do conceito de entidade/tabela, pode ser visto como uma categoria de objetos ou coisas que existem no mundo real (sujeito numa frase), sendo que esta categoria corresponde a um modelo representativo dos objetos desse tipo segundo Elmasri e Navathe, citado por Alalwan (2011). A instanciação de objetos concretos depende do modelo, sendo que como tal a entidade é somente o modelo deste objeto de acordo com o seu tipo. As entidades podem ser fortes (depende dela própria) ou fraca (depende de outras entidades, como o caso das entidades associativas ou com uma chave dependente). É o caso de cliente e contas, sendo que um cliente pode ter várias contas. Como tal o cliente é uma entidade forte e as contas são entidades fracas.

No caso de atributos/campos, segundo Elmasri e Navathe, citado por Alalwan (2011) e Vilannueva-Rosales (2011), correspondem a propriedades descritivas de uma entidade, pois cada entidade tem uma lista de atributos, sendo que cada atributo tem um domínio de valores (e.g. inteiros, caracteres) e restrições específicas (e.g. suporte de valores nulos ou valores por defeito). Os atributos podem ser classificados da seguinte forma:

- Simples (não pode ser dividido) ou composto (pode ser dividido em atributos simples);
- Pode ser considerado derivado se for possível obter o mesmo valor a partir de atributos, como o caso de idade calculada a partir de data de nascimento;
- Pode ser um atributo chave quando um registo só tem um valor possível para o atributo, sendo que pode ser composto com outros atributos para forma a chave. Esta chave pode ser primária quando corresponde ao identificador natural de cada registo.

No caso de relações que podem ser implementadas como restrições em base de dados, Alalwan (2011) e Vilannueva-Rosales (2011) citam Elmasri e Navathe, além de Date, para descreverem a seguinte forma de relacionar entidades:

- Pela cardinalidade da relação, podem ser ligações unárias quando a entidade se relaciona consigo própria, relações binárias quando a relação é estabelecida entre duas entidades ou n-ária quando a relação envolve mais do que duas entidades;
- No caso de relação entre duas entidades, a cardinalidade pode ser classificada como um-para-um (uma instância só se relaciona com outra instância entre duas entidades), um-para-vários (uma instância de uma entidade tem relação com várias instâncias de outras entidades) ou vários-para-vários (várias instâncias de uma entidade têm relação com várias instâncias de outras entidades). Esta cardinalidade pode ter restrições estabelecidas em termos de mínimos e máximos nas relações.

Nas relações/restrições acima descritas, aparece o conceito de chave estrangeira, que corresponde a um atributo/campo colocado numa das tabelas da relação. Por exemplo, numa relação de “cliente” tem várias “contas”, na tabela de “contas” aparece o atributo/campo “id cliente” enquanto chave estrangeira para garantir a relação com a tabela/entidade “cliente”.

No processo de modelação da estrutura de dados, um dos aspetos críticos está relacionado com a normalização e que de acordo com Alalwan (2011) e Villanueva-Rosales (2011), citando Elmasri e Navathe, pode ser definido da seguinte forma:

- **Primeira forma normal:** cada atributo ou conjunto de atributos da entidade é atómico, isto é, não se pode repetir na entidade;
- **Segunda forma normal:** está na primeira forma normal e os atributos que não são chave estão dependentes da totalidade da chave. Isto é, os atributos só fazem sentido se tiverem alguma relação com chave que representa de forma unívoca cada instância;
- **Terceira forma normal:** está na segunda forma normal e nenhum atributo não chave pode depender da chave por intermédio de outro atributo.

Para implementar, gerir e utilizar uma base de dados torna-se necessário uma linguagem, que no caso de bases de dados relacionais é o SQL (*Structured Query Language*). Esta linguagem pode ser decomposta em quatro tipos:

- **DDL (*Data Definition Language*):** para criar e alterar as estruturas de dados em termos de criação, alteração e eliminação das estruturas. Para o efeito utilizam-se construtores como CREATE, ALTER e DROP que se podem aplicar a tabelas, índices ou *views* de tabelas;
- **DML (*Data Manipulation Language*):** para interagir com as estruturas em termos de consulta, alteração e eliminação de instâncias (isto é, registos) de entidades. Para o efeito utiliza-se uma estrutura base de SELECT <lista de atributos> FROM <tabelas> WHERE <restrições> GROUP BY <lista de atributos de agregação> HAVING <restrições sobre agregações> e ORDER BY <lista de atributos para ordenação>. Em adicional pode-se utilizar ainda construtores de UNION (unir o resultado de duas consultas) e DISTINCT (aplicada a atributos em SELECT para considerar somente instâncias distintas). Além da consulta via SELECT, pode-se ainda utilizar construtores para INSERT (inserir dados em tabelas), UPDATE (atualizar dados em tabelas) ou DELETE (eliminar dados em tabelas);

- **DCL (Data Control Language):** linguagem com construtores para segurança como GRANT e REVOKE de permissões de utilizadores a estruturas de dados;
- **DTL (Data Transaction Language):** linguagem para suporte a lógica de transações, recorrendo a construtores de COMMIT (fechar a transação), ROLLBACK (reverter o que a transação efetuou) ou BEGIN WORK (marcar o início de uma transação).

A linguagem SQL apesar de ter uma estrutura base, tem implementações específicas por fabricantes de sistemas de gestão de bases de dados. No caso de bases de dados informacionais, pode-se utilizar o SQL, com ênfase para cláusulas de *Select*, *Where*, *Group By*, *Order by* e *Having*, sendo que alguns fabricantes como a Microsoft utilizam uma linguagem designada MDX (*Multidimensional Expression*) para trabalhar com soluções analíticas suportadas em *Microsoft Analysis Services* que é um motor OLAP da Microsoft. Considerando uma relação entre bases de dados relacionais e bases de dados semânticas (bases de dados definidas com modelos de dados de ontologias, como o OWL), Astrova (2009) considera um modelo de ontologia apresentado na Figura 3.22.

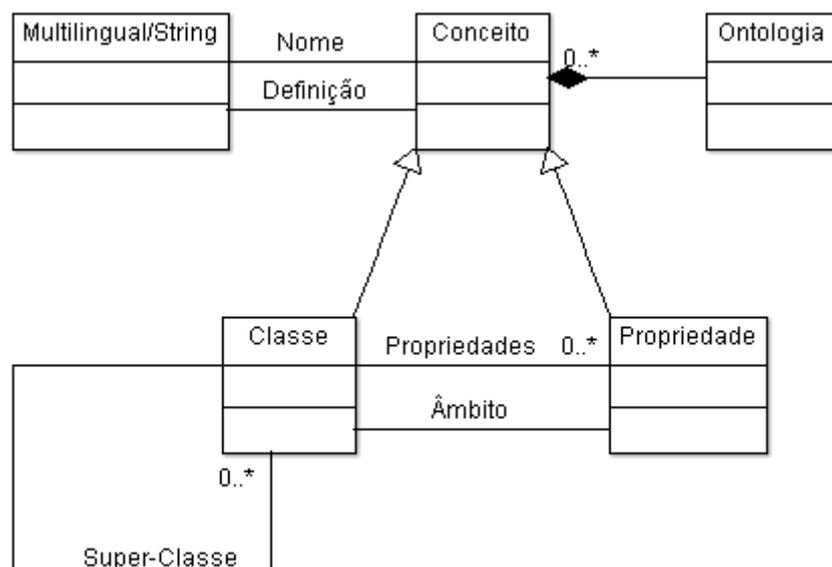


Figura 3.22: Modelo de uma ontologia, adaptado de Astrova (2009)

Tendo por base este modelo, Astrova (2009) define um conjunto de regras de correspondência, de onde se destacam as seguintes:

- **Entidade:** em termos de modelo de dados físicos é uma tabela. Corresponde a uma class, exceto se todas as suas colunas/atributos forem chaves estrangeiras, caso em que cada coluna/atributo corresponde a object property;
- **Atributos:** em termos de modelo de dados físico é um campo. Corresponde a data property, sendo que os domínios de valores correspondem a formatos XSD na ontologia. No entanto, caso sejam chaves estrangeiras, são consideradas como object property. No caso de ter uma restrição de campo como chave primária, corresponde a inverse functional property;
- **Restrições:** em termos de modelo de dados físico é forma de restrição. Corresponde a chaves estrangeiras, que pode ser implementada como object property, sub-class, symmetric property ou transitive property.

Villanueva-Rosales (2011), considera um modelo de correspondência muito equivalente, além de considerar que os registos numa tabela correspondem a Individual, de acordo com o modelo definido da ontologia, tal como apresentado na Tabela 3.3.

Modelo entidade-associação	Modelo relacional	Modelo OWL
Entidade	Tabela	Classe
Relação binária	Tabela ou campo	Propriedade de objetos
Relação N-ária	Tabela	Não disponível
Especialização	Tabela e integridade referencial	Hierarquia com Sub-classe Hierarquia com “parte-de”
Mapeamento de tipos de união	Campo	Classe
Instância	Linha ou registo	Indivíduos
Atributos (simples, compostos, multi-valor)	Campo	Propriedade de dados
Atributos chave	Chave primária	Propriedade de objetos de função inversa
Conjunto de valores	Domínio	Domínio
Cardinalidade	Chaves	Restrição de cardinalidade
Referência cruzada	Integridade referencial e gatilhos	Restrição de cardinalidade e de valor

Tabela 3.3: Modelo de relação entre base de dados relacional e ontologia, adaptado de Villanueva-Rosales (2011)

3.4 Sistemas de informação

3.4.1 Sistemas de informação

Os sistemas de informação são compostos por uma arquitetura de aplicações, informação e tecnologia de suporte, além de processos de gestão para garantir que atingem o objetivo de gerir e disponibilizar a informação com qualidade, desempenho, segurança e no contexto das necessidades de negócio. São como tal uma das componentes do modelo de negócio, sendo representadas ao nível da arquitetura empresarial. Segundo Potter et al. (2005), os sistemas de informação permitem recolher, processar, armazenar e disseminar informação, sendo que como qualquer sistema, utilizam entradas (dados) e saídas (cálculos e relatórios), operam dentro de um ambiente e incluem mecanismos de controlo. A visão sistémica em termos de fluxos de entrada, controlo e saída no processamento da informação é igualmente abordada por Laudon e Laudon (2012), tornando mais claro a relação com a organização em termos de ambiente externo e posicionando o sistema de informação numa relação sistémica entre ambiente interno (organização) e ambiente externo (ambiente) de acordo com o modelo apresentado na Figura 3.23.

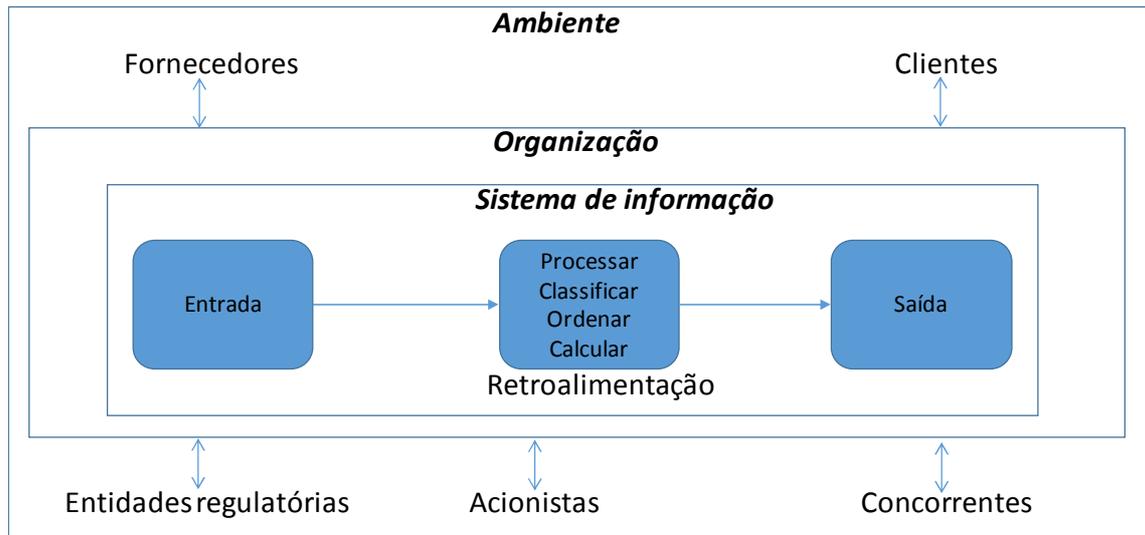


Figura 3.23: Fluxo de processamento informação, adaptado de Laudon e Laudon (2012).

De acordo com Laudon e Laudon (2012), os sistemas de informação correspondem assim a um conjunto de componentes interligados que permitem recolher, processar, armazenar e distribuir a informação para tomada de decisão e controlo da própria organização. Para Keri et al. (2006), os sistemas de informação combinam tecnologia, pessoas e processos para facilitar a comunicação de dados, informação e conhecimento.

Em termos de tipologia de sistemas, Laudon e Laudon (2000) definem os sistemas em níveis estratégicos, gestão, conhecimento e operacional, sendo que cada nível tem um tipo de utilizador específico e tratam uma parte de áreas funcionais da organização, como apresentado na Figura 3.24.

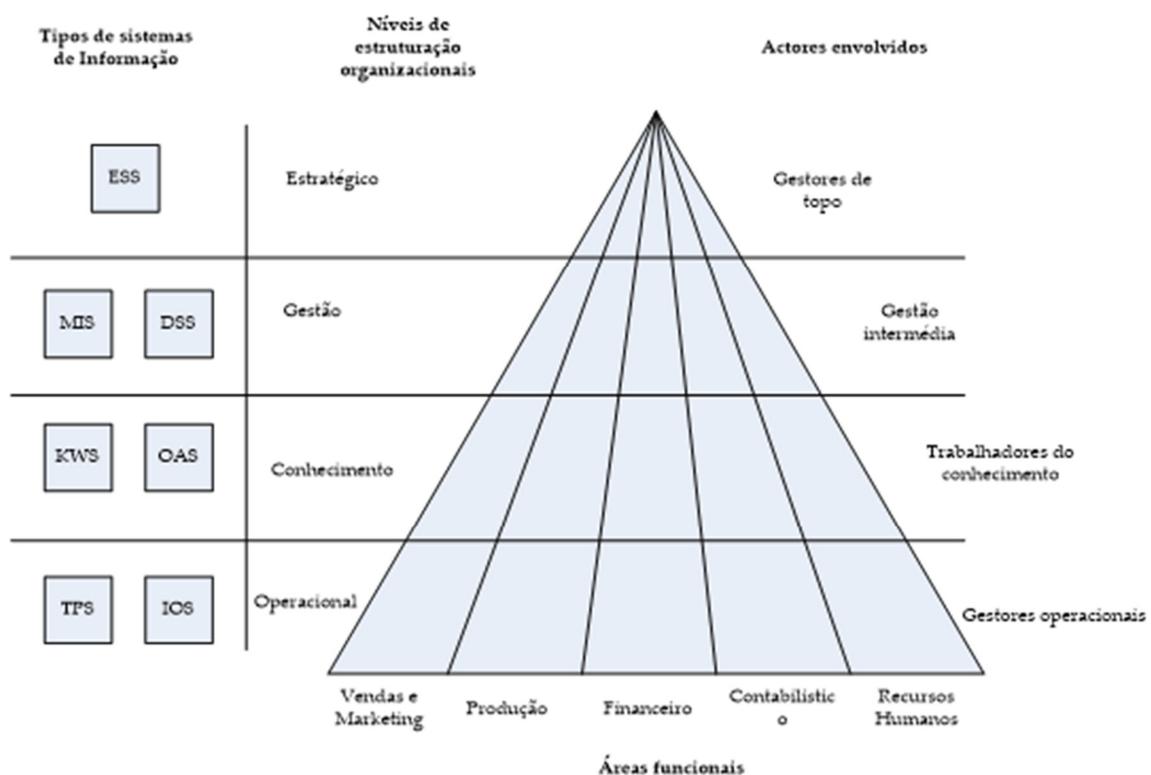


Figura 3.24: Tipologia de sistemas de informação segundo Laudon e Laudon (2000)

As áreas funcionais correspondem a funções da organização, como parte do modelo de negócio da organização, sendo que cada área tem um conjunto de processos de negócio. Por outro lado, os níveis de estruturação organizacional para agregar tipo de sistemas de informação permitem criar uma separação em função da produção e consumo de informação em torno de sistemas operacionais (apoio às operações) e sistemas informacionais (apoio à gestão), de acordo com O'Brien (2004):

- **Sistemas operacionais:** segundo O'Brien (2004), estes sistemas suportam o processamento de dados gerados pelas operações realizadas pela organização referente a transações e colaboração entre equipas de trabalho. Pelo modelo de Laudon e Laudon (2000), correspondem aos sistemas de conhecimento e sistemas operacionais que produzem informação no contexto de processos de negócio de acordo com as áreas funcionais. Normalmente cada área funcional tem o seu conjunto próprio de informação. Podem produzir informação de gestão, mas normalmente para orientação tática. Dependem normalmente de dados em real-time e têm um grande volume de utilizadores de vários tipos de acordo com a estrutura orgânica;
- **Sistemas informacionais:** para O'Brien (2004), estes sistemas estão baseados em formas de relatórios para disponibilizar informação para a tomada de decisão. Pelo modelo de Laudon e Laudon (2000) correspondem aos sistemas de gestão e sistemas estratégicos. Dependem da informação produzida pelos sistemas operacionais, criando modelos de análise dos padrões desta informação para tomada de decisão quer de forma vertical em cada área funcional, quer de forma transversal mais alinhada com a estratégia, agregando informação de várias áreas funcionais. À informação recolhida dos sistemas operacionais, acrescenta-se informação de contexto estratégico relativo a concorrentes, clientes e economia dos locais onde têm atividades, para suporte à decisão relativa a ambiente externo estratégico. Podem necessitar de informação em *real-time*, mas para monitorização em ciclos de gestão, agregam informação para análise em periodicidades semanais, mensais, trimestrais, semestrais e anuais. Têm um número menor de utilizadores, que correspondem normalmente à direção da empresa, quer a níveis intermédios quer ao nível de topo.

Esta separação entre tipo de sistemas resulta igualmente da análise de várias perspetivas como olhamos para a questões críticas para sistemas de informação, como é o caso de dados (captura e armazenamento, tipo de modelo, dimensão), aplicações informáticas (aplicações Web ou Desktop), tipo de transações (concorrência, tempo da transação), forma de entrada de dados e quantidade de utilizadores. No caso da perspetiva de análise de dados, destaca-se a captura e armazenamento pelo aspeto crítico da dimensão temporal associada a cada informação na base de dados em cada tipo de sistema (e.g. real-time para captura associada à ocorrência dos processos, ou captura de dados com histórico e mantendo esse histórico na própria base de dados, gerado o que é designado por fotografias diárias de posições da informação). Considerando uma análise por estas perspetivas Mawilmada (2011) cita Kadlec ao considerar que os sistemas operacionais são sistemas do tipo OLTP (*OnLine Transaction Processing*) para suporte a processos operacionais das organizações enquanto os sistemas informacionais são sistemas do tipo *Data Warehouse* para suporte a informação estratégica e tática das organizações, detalhando a comparação entre várias perspetivas de acordo com o que é apresentado Tabela 3.4.

Perspetiva	OLTP	Data Warehouse
Objetivo	Suporta as operações de negócio do dia-a-dia	Suporta a tomada de decisão a nível tático e estratégico
Utilizadores	Grande quantidade de utilizadores	Tradicionalmente com pequena quantidade de utilizadores, mas com tendência para expansão para mais utilizadores com acessos a dados especializados
Captura e armazenamento de dados	Dados em real-time geridos em aplicações de negócio como ERP	Dados históricos, obtidos e tratados a partir de sistemas OLTP com o objetivo de se efetuar análises e relatórios sobre tendências e padrões críticos para a organização.
Dimensão da base de dados	Pequena e média dimensão, tipicamente entre 1 a 100 gigabytes	Grandes bases de dados entre 100 gigabytes e vários terabytes que são obtidos a partir de sistemas OLTP, colocados em servidores de <i>datastage</i> onde são agregados e transformados antes do carregamento final na <i>data warehouse</i>
Desenho do modelo de dados	Modelo relacional normalizado	Modelo em <i>star</i> ou <i>snowflake</i> com pré-agregação
Aplicações Front-end	Aplicações web ou desktop com entrada de dados em caixas de texto e funcionalidades associados aos ciclos de negócio	Ferramentas analíticas com visualização gráfica e com capacidade para agregar e detalhar visões de dados até ao detalhe
Transações concorrentes	Número elevado de transações em simultâneo	Número baixo de transações em simultâneo, mas com tendência para crescimento
Tempo por transação	Tempo curto ao nível de milissegundos com pequena quantidade de dados a retornar ao utilizador	Tempo médio e elevado com utilização de funções de agregação (e.g. SUM, AVG, STDEV)
Entrada de dados	A partir de utilizadores	Dados migrados a partir de sistemas OLTP

Tabela 3.4: Sistemas operacionais vs sistemas informacionais, adaptado de Mawilmada (2011)

3.4.2 Arquitetura de sistemas de informação

Segundo Potter et al. (2005) os sistemas de informação utilizam tecnologia de informação e telecomunicações enquanto componentes do sistema, sistematizando como componentes o *hardware*, *software*, bases de dados e redes de comunicações, mas também os procedimentos e as pessoas. Na mesma linha de pensamento, Laudon e Laudon (2004) consideram que existem vários componentes (*hardware*, bases de dados e redes de comunicações) como base para o software criar uma interdependência com a estratégia da empresa e seus processos de negócio, como apresentado na Figura 3.25.

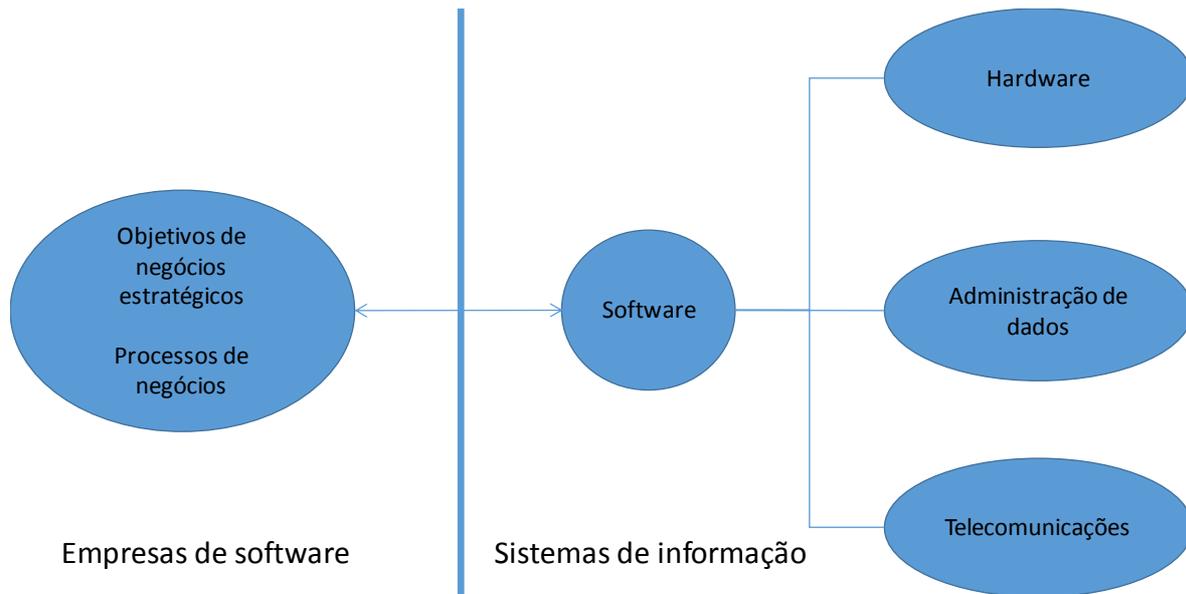


Figura 3.25: Relação entre Organização e Sistemas de Informação, adaptado de Laudon e Laudon (2004)

Esta separação entre componentes de sistemas de informação apresentado na Figura 3.25, é abordada no contexto da arquitetura empresarial ao se considerar a organização representada na arquitetura de negócio e os sistemas de informação na arquitetura aplicacional, arquitetura de informação e arquitetura tecnológica. De acordo com alguns dos referenciais abordados no sub-capítulo 3.3.3.1, temos o seguinte:

- Para Zachman (2009) os sistemas de informação correspondem às dimensões *HOW* e *WHEN*;
- Para ArchiMate, segundo GROUP (2012, 2016), os sistemas de informação são representados pelo *application layer* e *technology layer*;
- TOGAF (2013) refere que os sistemas de informação são representados ao nível do *information system architecture* e *technology architecture*.

De notar, no entanto, que a separação específica da arquitetura de informação, apesar de ser uma preocupação ao nível geral de arquitetura empresarial e ao nível de sistemas de informação, não é normalmente representada de uma forma explícita. Este é um dos aspetos fundamentais ao nível do problema e da hipótese de solução abordada na nossa investigação.

3.4.3 Sistemas informacionais

3.4.3.1 Definição

Um dos tipos de sistemas identificados por Laudon e Laudon (2012) e O'Brien (2004), corresponde ao que é designado genericamente como sistemas de informação de gestão, que na tese consideramos como *business intelligence*. Laudon e Laudon (2012) consideram como sistemas informacionais o conjunto de sistemas MIS (*Management Information System*), DSS (*Decision Support Systems*) e EIS (*Executive Information System*).

O termo *business intelligence* foi criado nos anos 80 pelo Gartner tendo várias aproximações de definições que são comuns na inclusão de extraírem dados a partir de sistemas operacionais em várias fontes e formatos para transformar, limpar os dados e armazená-los em estruturas de dados informacionais para posterior análise por ferramentas específicas para focar em informação estratégica e tática, como referidos por vários autores:

- Para Barbieri (2001) corresponde a um tipo de sistema que integra múltiplas fontes de informação para se definir a estratégia e formas de atuação da empresa, permitindo aos gestores definir metas e formas de as alcançar;
- Arnet et al. (2000) e Moss e Altre (2003) concordam que é um tipo de sistema que permite combinar dados, informação e conhecimento acerca do ambiente interno e externo do negócio, com o objetivo de suportar a tomada de decisão, facilitando assim a gestão da dinâmica da organização e das suas vantagens competitivas;
- Rossini e Palmisano (2003) referem que estes sistemas combinam conceitos de *data warehouse*, *data mart*, OLAP, com sistemas específicos de EIS, *data mining* e DSS;
- Para Gangadharan e Sundaravalli (2004) este tipo de sistemas corresponde a uma arquitetura que tem como objetivo a integração de vários componentes, bases de dados e aplicações operacionais, consolidadas para dar acesso a dados organizados com sentido adequado para vários níveis de tomada de decisão consistentes no contexto do desempenho e monitorização efetiva da estratégia e operações. Para este autor, o *business intelligence* representa assim uma nova disciplina de gestão em que os dados são tratados finalmente como recursos empresariais.

Para Kung et al. (2005), na mesma linha de pensamento de Gangadharan e Sundaravalli (2004), este tipo de sistemas pode ser uma solução para um problema de monitorização de desempenho organizacional. Este autor refere que muitas organizações com dimensão (o estudo é sobre o Credit Suisse), enfrentam a complexidade e heterogeneidade da sua arquitetura de sistemas de informação, enquanto integração de várias aplicações, tecnologias e histórico de implementação, incluindo a utilização de *workflow* de suporte ao negócio que capturam só parte dos processos de negócio e eventos associados. Na sua investigação sobre alternativas para suporte à monitorização e medição de processos, Kung et al. (2005) destaca a monitorização geral de atividades de negócio, designado por BAM (*Business Activity Monitoring*), que é um sistema que monitoriza em *real-time* os indicadores de desempenho de processos de negócio. No entanto, é parte integrante da exploração em *business intelligence*, pois a sua implementação recorre normalmente a *data warehouse*, como indicado pelo autor como uma das formas de implementação.

Este tipo de sistema informacional deve ser visto pela necessidade de integrar informação para o ciclo base de gestão enquanto funções de planear, dirigir, monitorizar, avaliar e reportar o estado da organização, agregando dados em informação e esta em conhecimento, com inteligência associada à análise mediante posicionamento de alertas para desvios, baseado em conceitos de métricas (indicadores) e dimensões (perspetivas de análise). No entanto face à heterogeneidade de dados envolvidos, o fator crítico de sucesso está em encontrar uma forma de dar sentido e integrar de forma lógica esses dados, de várias fontes e formatos, aplicando-se as ferramentas e técnicas adequadas para a sua análise, organização e para guiar o suporte à decisão com base nessa informação assim estruturada, que é um tema de metadados investigado nesta tese.

3.4.3.2 Arquitetura de componentes *business intelligence*

Laudon e Laudon (2012) ao abordarem os sistemas de apoio à decisão considera uma arquitetura base dos sistemas de *business intelligence*, tal como apresentado na Figura 3.26, enquanto sistema que a partir de transações internas e dados externos, permite agregar informação, numa lógica de transformação de dados em informação, para servir de base para sistemas de apoio à decisão. De notar nesta representação, o diretório de informação como génese do conceito de metadados, apesar de estar somente relacionado com as estruturas de dados e não com a exploração e com a transformação.

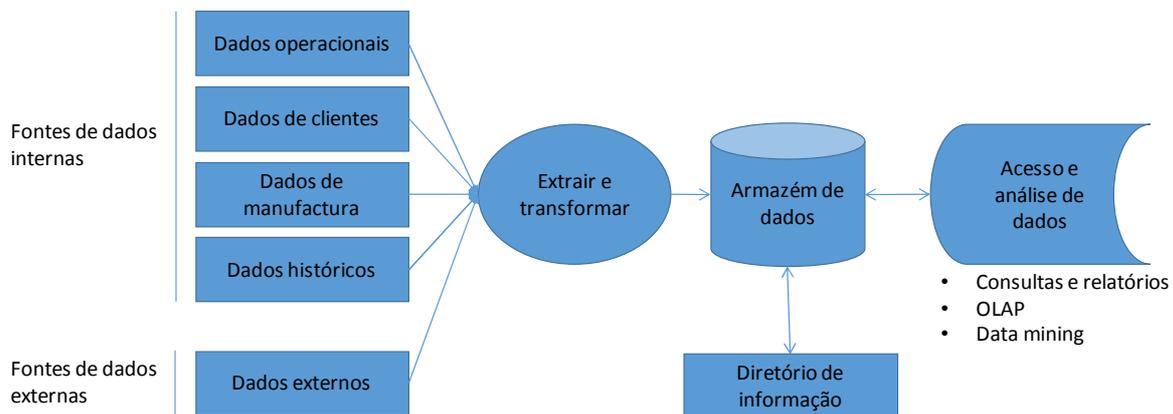


Figura 3.26: Componentes *data warehouse*, adaptado de Laudon e Laudon (2012)

Kimbal e Ross (2013) focando na componente mais dimensional de organização de dados em termos de *data mart*, apresentam uma arquitetura como representado na Figura 3.27, onde referem o mesmo fluxo de transformação (via ETL), armazenamento (via Enterprise DW) e exploração via modelos analíticos (via *Star Schema* e OLAP, ou *BI Application*). Mas consideram sempre a base da informação a partir do *source transaction*, que são os sistemas operacionais.

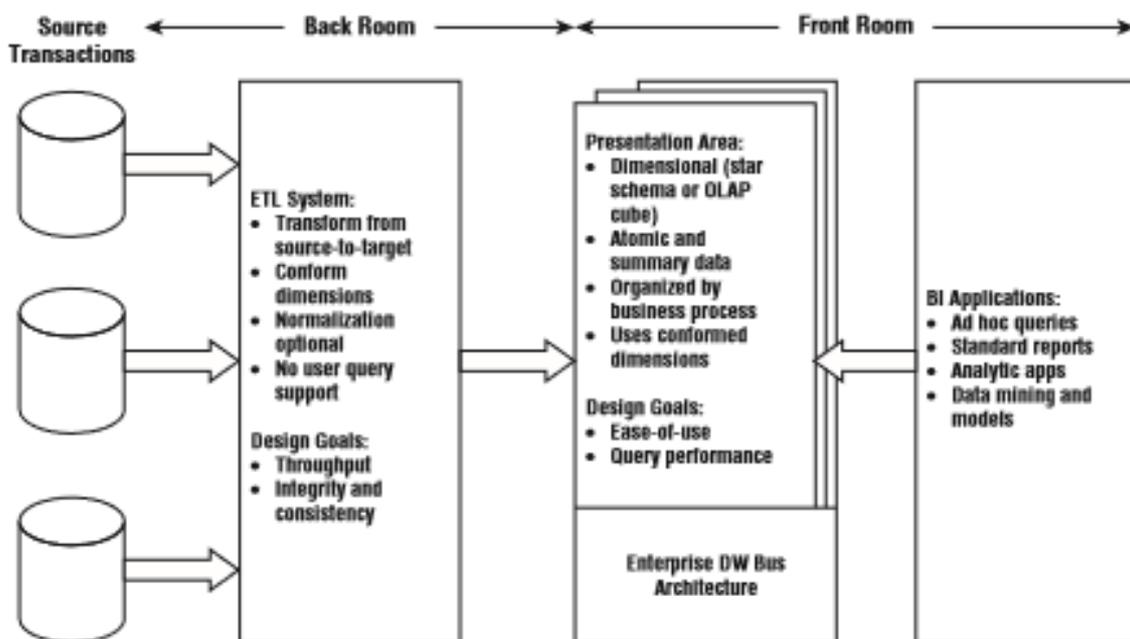


Figura 3.27: Componentes *data warehouse* (Kimbal e Ross, 2013)

Para Marco (2013) estes sistemas permitem o carregamento de bases de dados designadas ODS para efeito de *reporting* tático (ao nível de sistemas operacionais), além de permitirem carregar estruturas de dados designadas por *data warehouse* e *data mart*, sendo existe uma dependência entre *data mart* e *data warehouse*, tal como apresentado na Figura 3.28. O ODS tem dados mais em bruto (designado por *raw data*) e temporária, enquanto a *data warehouse*, tem uma noção de histórico com dados mais permanentes no tempo, como base para agregações temáticas funcionais ao nível de *data mart* (normalmente em modelos *star* ou *snowflake*).

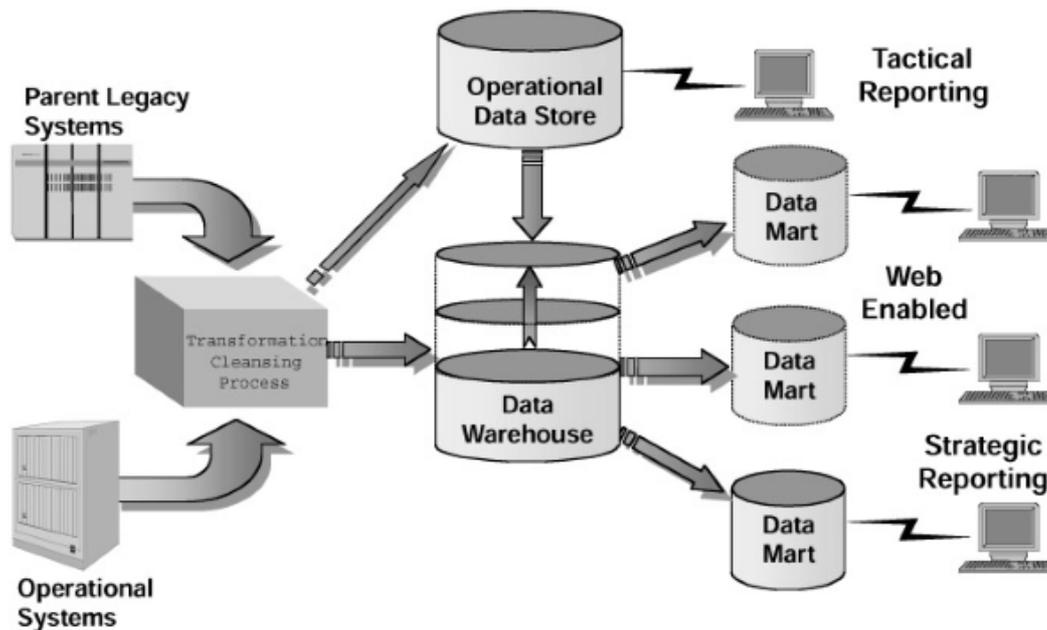


Figura 3.28: Data Warehouse Data Architecture (Marco, 2013)

Para Chaudhuri et al. (2011), tendo por base fontes de dados ao nível de aplicações operacionais ou dados externos, os sistemas de *business intelligence* têm uma arquitetura decomposta em quatro níveis (ETL como *Data Movement*, *Data WareHouse Server*, *Mid-tier Server* como *Data Mart* e *Front-End Application* para exploração) como apresentado na Figura 3.29. Não considera metadados, mas introduz o conceito de MapReduce Engine para dados em BigData ao nível do *data warehouse server*. No entanto, a arquitetura está alinhada com os autores anteriores.

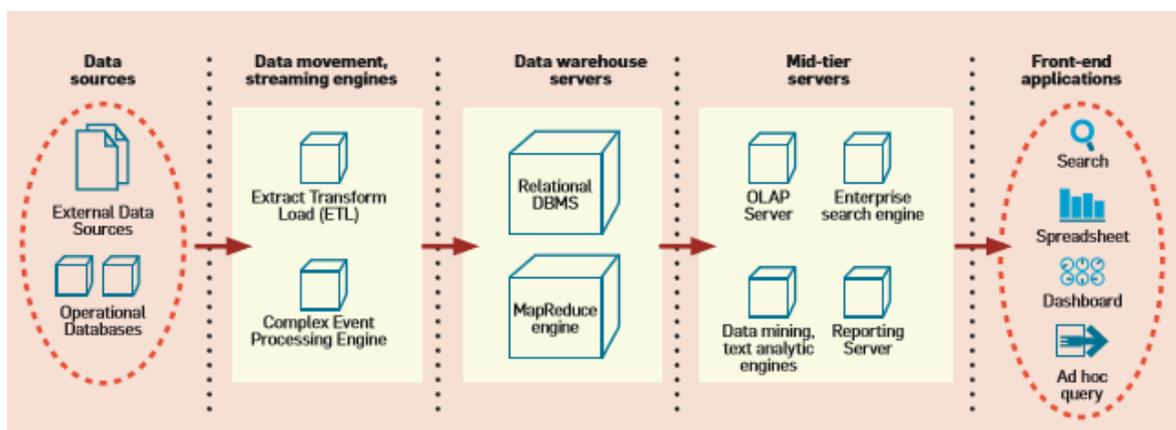


Figura 3.29: BI Architecture (Chaudhuri et al., 2011)

Ao nível da exploração da informação de gestão previamente integrada e armazenada em dados informacionais o modelo de *business analytics framework* da Gartner separa as necessidades de utilização, produção e disponibilização de informação de gestão, separando ao nível de processos analíticos as perspetivas de *descriptive* (o que aconteceu), *diagnostic* (porque aconteceu), *predictive* (o que poderá acontecer) e *prescriptive* (o que posso fazer para otimizar os objetivos), tal como apresentado na Figura 3.30.

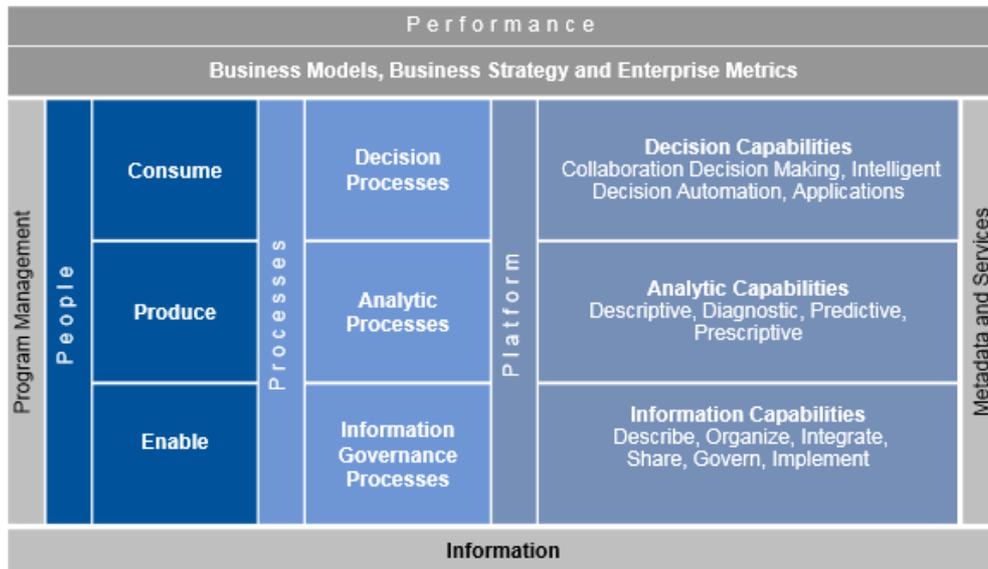


Figura 3.30: Gartner business analytic framework (Gartner, 2011)

A arquitetura de um sistema de *business intelligence* pode assim ser resumida de acordo com a representação na Figura 3.31, com a particularidade de separação entre componentes de integração, dados e acessos, posicionando os metadados de forma transversal à arquitetura de sistemas de informação operacional e informacional.

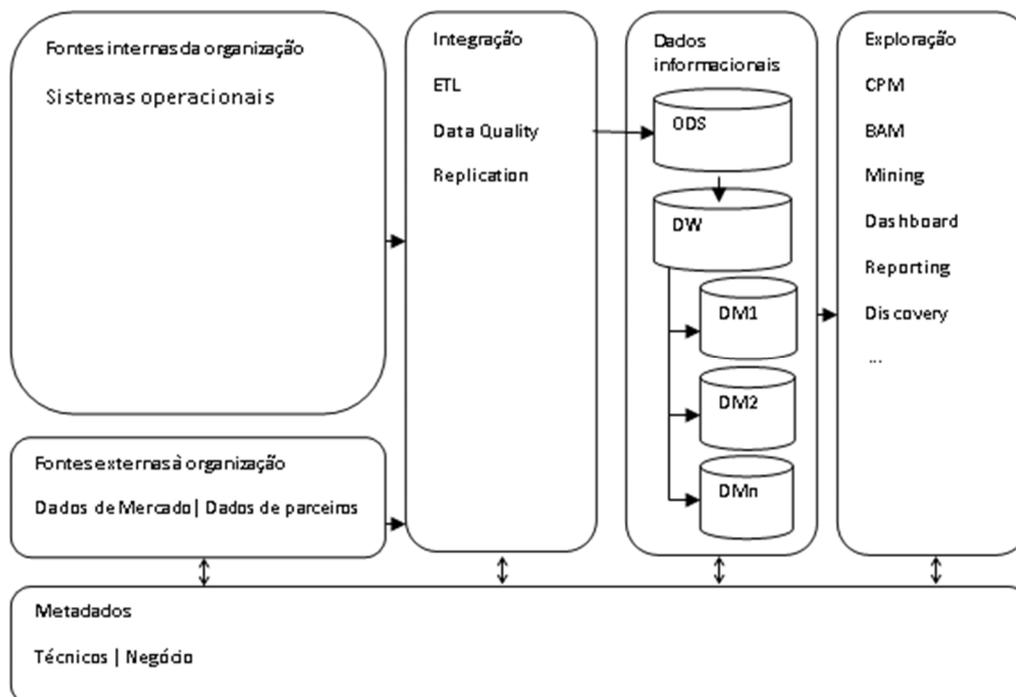


Figura 3.31: Arquitetura consolidada de *business intelligence*

Esta arquitetura considera os seguintes componentes

- **Integração de dados:** processos de mapeamento entre fontes e destinos, sendo que as fontes podem ser operacionais, ou fontes da própria *data warehouse* em termos de estruturas de dados. Estes processos permitem agregar, transformar, limpar e calcular campos adicionais, carregando em estruturas de dados para utilização, ou para registo de controlo de qualidade de dados. Pode-se recorrer a ferramentas de ETL (*Extraction, Transform and Load*) ou *data replication* (utilizam mecanismos de log para detetar alterações em bases de dados), além de ferramentas específicas para identificar e corrigir problemas de qualidade de dados;
- **Armazenamento de dados:** modelos de dados com vários fins que correspondem ao resultado dos processos de integração de dados e são a base para a exploração. Podem ser modelos de ODS (*Operational Data Store*) com uma visão de dados ainda perto dos sistemas operacionais, modelos de DW (*data warehouse*) com algumas transformações e agregações e modelos mais agregados e organizados por temas de análise ao nível de DM (*data mart*). Pode ainda integrar modelos de MapaReduce para BigData;
- **Exploração de dados:** estes processos incluem *data mining*, CPM (*Corporate Performance Management*), BAM (*Business Activity Monitoring*), *dashboard/scorecard* e *reporting ad-hoc*, entre outros. Nos processos dependentes de navegação, recorrem a processamento via OLAP, que pode ser multidimensional ou relacional, numa conjugação de definições técnicas entre estruturas de dados e processos de acesso e visualização dos dados;
- **Metadados:** modelo de dados com informação sobre estruturas de dados e regras utilizadas em todos os processos de integração e exploração. Normalmente, cada fabricante tem os seus modelos e componentes específicos de metadados.

3.4.3.3 Arquitetura de componentes de dados em *business intelligence*

Do ponto de vista técnico, Kimbal e Ross (2013) focam o conceito na componente de *data warehouse*, referindo-se de forma alargada a todo o componente de dados informacionais, enquanto forma de organização da informação, caracterizado da seguinte forma:

- Está organizado por assuntos que correspondem a áreas de negócio da organização e devem ser considerados para modelação quer ao nível de modelos base de carregamento como ODS e DW, e especificamente ao nível de *data marts*;
- Deve ser integrado, independentemente da origem dos dados e seus formatos, deve-se utilizar o processo de ETL ou *Data Replication*, para transformar e consolidar os dados, para carregar de forma integrada e sem duplicações;
- Mantém dados não voláteis e com uma data de manutenção superior a cinco anos, pois por um lado os dados são carregados normalmente por períodos e não dependentes de transações *online*, e por outro lado, necessitam de perspetivas históricas de acordo com os temas, para detetar padrões, tendências e comparações.

Segundo Berkeley (2006), é necessário considerar componentes de arquitetura próprios destes sistemas, como é o caso de “*data*”, “*application*”, “*technology*” e “*processes*”, tal como apresentado na Figura 3.32. Na componente específica de “*Data*”, Berkeley (2006) considera que uma *data warehouse* integra igualmente uma visão de projeções futuras em termos de previsões e orçamentos, o que permite montar uma estratégia de acompanhamento e medição entre previsto e realizado, de acordo com o planeado, identificando e planeando as melhorias necessárias face às evidências detetadas.

Data Warehouse Components

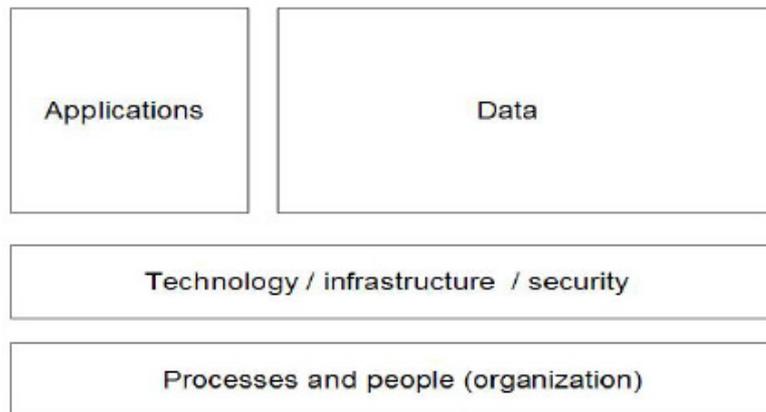


Figura 3.32: Componentes arquiteturais de uma *data warehouse* (Berkeley, 2006)

Como tal, para Berkeley (2006), uma *data warehouse* e *data mart* incluem dados históricos e uma separação entre dados mensuráveis (factos) e contexto ou perspetivas de análise (dimensões). As dimensões podem ser utilizadas em vários factos ocorridos, sendo que o conjunto de dimensões partilhadas representam o nível de visão comum das várias áreas de negócio estruturadas nos factos analisados sob estas perspetivas. O conhecimento destas estruturas, envolve um modelo de metadados enquanto guia ou dicionário sobre estruturas e utilização dos dados em termos técnicos e de negócio.

Considerando os vários níveis de dados nestes sistemas, podemos sistematizar da seguinte forma a função de cada um em termos de utilização e forma de modelação:

- **Operational Data Store e Data Staging:** são áreas de dados temporárias, apesar de poderem manter-se durante períodos superiores a um mês, mas que têm como objetivo suportar os processos de ETL na agregação, transformação e limpeza de dados, antes do carregamento otimizado para a *data warehouse* e/ou *data marts*. São dados em bruto, obtidos do operacional, normalmente com modelação relacional se for em bases de dados, ou podem ser dados não estruturado ou mesmo ficheiros diversos;
- **Data Warehouse:** modelos de dados organizados por temas, de acordo com o objetivo e negócio da organização, para suportar processos de *data mining*, CPM, BAM e visualização através de ferramentas de navegação utilizando normalmente tecnologia OLAP (*On-line Analytic Processing*). Podem utilizar *surrogate keys* (chaves ou códigos gerados ao nível do DW para normalizar casos em que vários sistemas origem têm o mesmo código, mas com definições diferentes), modelação com SCD (*Slowly Changing Dimension*), modelos em *star Schema*, *snowflake schema* ou totalmente relacional;

- **Data Mart:** modelo de dados organizado por assuntos normalmente associado a temas (e.g. rentabilidade, clientes, produtos, eficiência), dependente ou independente da *data warehouse*, com dados replicados, mas agregados de forma otimizada e com cálculos adicionais ao que existe na *data warehouse*, para acelerar a visualização, agregar conteúdos temáticos e facilitar a sua distribuição/acesso. É neste nível de dados que o conceito de métricas e dimensões de análise é mais utilizado. Utiliza normalmente modelação em *star schema* ou *snowflake schema*;
- **Metadados:** estrutura de dados que permitem identificar as fontes, manter informação sobre as transformações e identificar a relação entre origem e destino dos dados, suportando igualmente o processo de ETL e os modelos de visualização de dados. Permitem igualmente adicionar informação numa perspetiva de negócio, incluindo a explicação de fórmulas de cálculo, utilização de métricas e relatórios, ou inclusive a relação entre conceitos mapeados em campos de dados.

Na base dos modelos, as tabelas são normalmente definidas como sendo de Factos ou Dimensões segundo Kimbal e Ross (2013), o mesmo acontecendo para os campos das tabelas, mas sob o nome de métricas e dimensões. De acordo com Mohania et al. (1999), existe uma diferença entre métricas e dimensões:

- Métricas correspondem a dados normalmente numérico ou passível de ser factual (por exemplo contagens de registos com determinadas características) que representam a atividade da organização (e.g. quantidade de vendas, valor das vendas, quantidade de clientes, rentabilidade de produtos);
- Dimensões correspondem a perspetivas de análise das métricas, isto é, da atividade da organização (e.g. clientes, regiões, semestres, produto).

Estas métricas e dimensões, são organizadas em modelos específicos com tabelas que podem ser de factos ou de dimensões, estando organizadas num modelo em *star schema* ou *snowflake schema*, sendo a base do que é conhecido como bases de dados multidimensionais, segundo Chaudhuri et al. (2011) ou também bases de dados informacionais.

No caso de formato em *Star*, as tabelas de factos têm um conjunto de métricas, designadas medidas de análise utilizando-se uma chave primária composta por várias dimensões, que são igualmente chaves estrangeiras para as tabelas e dimensões recorrendo a Kimbal e Ross (2013), como apresentado na Figura 3.33. Nesta modelo considera-se sempre um grupo de uma tabela de factos e várias tabelas de dimensões, mas as tabelas de dimensões não têm relações entre si. As hierarquias de dimensões (e.g. tipo produto, produto, categoria, para o caso de "*Product Dimension*") estão representados numa única tabela de dimensão de forma desnormalizada. Este modelo apresenta ganhos no desempenho, pois a *query* de exploração junta normalmente a tabela de factos com tabelas de dimensões necessárias, mas sem decomposição das dimensões em várias outras junções pois a hierarquia está toda representada numa única tabela.

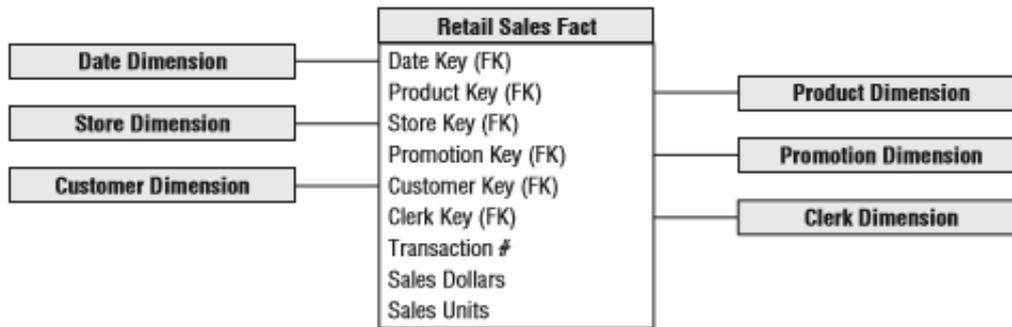


Figura 3.33: Modelo multidimensional star schema (Kimbal e Ross, 2013)

No caso de formato em *snowflake schema* as tabelas de dimensões estão normalizadas de onde resulta a criação de hierarquias em dimensões, como apresentado na Figura 3.34 para o caso desagregado da dimensão “Product Dimension”. Desta forma, para se obter os mesmos dados que o modelo em *Star*, obriga a juntar mais tabela numa *query*, o que pode prejudicar o desempenho.

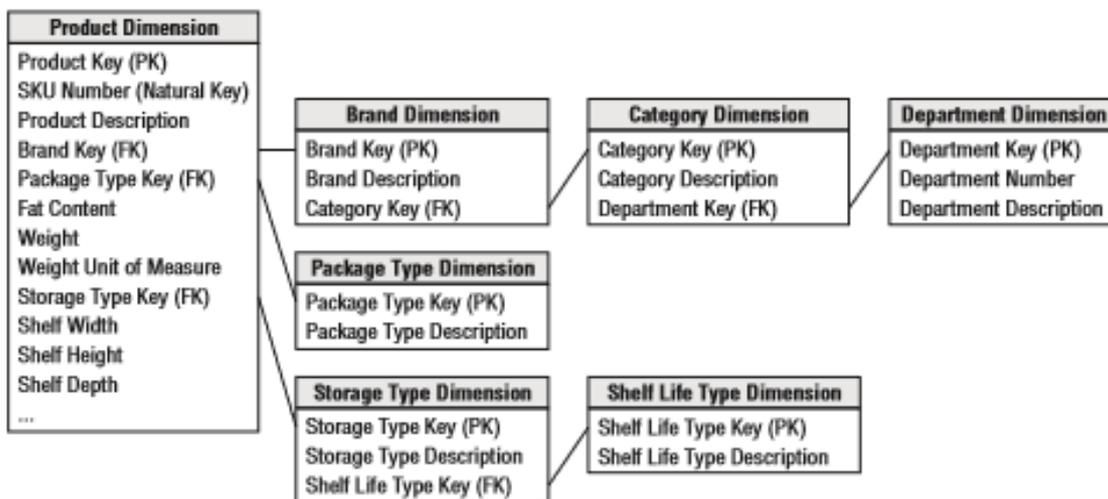


Figura 3.34: Modelo multidimensional em snowflake (Kimbal e Ross, 2013)

Por outro lado, ao nível de *data warehouse* e respetiva relação com *data marts*, existem metamodelos sectoriais desenvolvidos por fabricantes com modelos com os principais conceitos (tabelas, campos, relações), que podem ser implementados pelas organizações com os devidos ajustamentos face ao tipo de dados e modelo de negócio específico de cada organização, como é o caso do *IBM Banking Data Warehouse Blueprint* [IBM BDW] (2014) e *Teradata Financial Services Data Model* [TERADATA] (2014).

3.4.3.4 Metodologias *business intelligence*

Segundo Gofarelli e Rizzi (2009) existem duas abordagens para implementação de sistemas de *business intelligence*:

- **Top-Dow:** abordagem em cascata de análise de requisitos, desenho e implementação de todo o sistema. É uma abordagem com custos normalmente elevados e que torna clara a dificuldade de obter todos os requisitos e relacionar os requisitos com fontes de dados e regras de transformação, além de ajustamentos de necessidades ao nível da exploração de dados;

- **Bottom-Up:** abordagem focada na criação dos dados e ciclos de implementação de exploração de dados. É a abordagem mais utilizada;

Ao nível de ferramentas, existe uma classificação das mesmas por vários analistas de tendências com destaque para a Gartner (2015) que apresenta modelos de Quadrantes Mágicos para o seguinte tipo de ferramentas:

- **Integração:** *data integration tools* e *data quality*, onde são consideradas as ferramentas de *ETL*, *data replication*, *metadata*, *data quality* e *data federation*;
- **Exploração:** *business intelligence*, *analytics plataform* e *advanced analytics* onde são consideradas as ferramentas de *dashboard*, *reporting*, *data discovery*, *self-service-BI* e *data mining*;
- **Appliances e bases de dados:** *Data wareHouse*, onde são consideradas bases de dados, *appliances* e soluções *BigData*.

3.4.4 Exploração de conhecimento

3.4.4.1 Definição

Knowledge & data discovery é um processo não trivial para identificar e entender padrões de dados que sejam válidos, novos e úteis, segundo Fayyad et al. (1996). Os principais objetivos do *knowledge discovery* sobre bases de dados são os seguintes, de acordo com Kuo et al. (2007):

- **Preditiva:** tendo por base alguns atributos, permitir prever valores e padrões futuros e desconhecidos;
- **Descritiva** permitir obter uma descrição dos dados em termos de padrões, mas sem recorrer a modelos preditivos, sendo executado diretamente pelo utilizador.

Para se implementar estes processos recorre-se a estatística, bases de dados e inteligência artificial, segundo Piatetsky-Shapiro e Matheus (1992), utilizando-se uma sequência de tratamento de dados para converter em conhecimento, como apresentado na Figura 3.35 de acordo com Fayad et al. (1996):

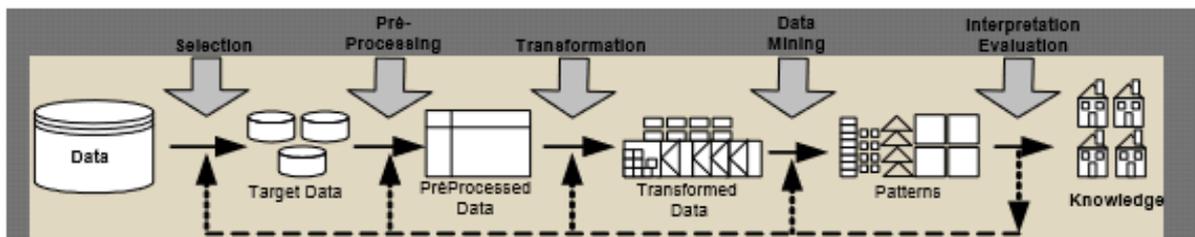


Figura 3.35: Processo knowledge discovery (Fayad et al., 1996)

De acordo com este processo, existem as seguintes etapas críticas, segundo Fayad et al. (1996):

- **Selection (seleção):** selecionar os dados relevantes de acordo com o tipo de problema a resolver;
- **Data pre-processing (pré-processamento de dados):** validar e efetuar limpezas de qualidade de dados, definindo valores por defeito, harmonização de conceitos para permitir relacionamentos, validações por cada atributo ou entre atributos relacionados;
- **Data transformation (transformação de dados):** transformar os dados para o nível de granularidade necessária para análise posterior, além de efetuar agregações que sejam necessárias;
- **Data mining (mineração de dados):** utilização de técnicas e modelos de *data mining* (e.g. *Statistical Regression, Decision Tree, Neural Network, Fuzzy Modelling*) de acordo com a natureza dos dados e tipo de problema a resolver;
- **Interpretation and evaluation (interpretação e avaliação):** validar os modelos e os resultados, incluindo atributos não utilizados e adequação à resolução do problema.

Para Buckinx et al. (2007), atualmente, derivado do grande volume de dados, necessidades de *Marketing* e novas soluções de *business intelligence* focadas em *data discovery* (Gartner, 2015), o conceito de *knowledge discovery* funde-se com o conceito de *data discovery* quer sobre os dados da organização, quer inclusive sobre a identificação de dados que podem existir (metadados).

Centrando na relação entre os metadados vistos como ontologia para processos de *knowledge & data discovery* existem soluções específicas que podem ser utilizadas, como é detalhado nos sub-capítulos seguintes e aqui resumido

- **Ontology learning:** integração e reutilização de ontologias de forma semiautomática;
- **Data lineage:** análise de causa-efeito e de impacto sobre estruturas de dados, para entender as regras de transformação de dados até ao conhecimento, normalmente aplicado em *business intelligence*, mas que tem hoje um potencial de reutilização de forma transversal nas organizações ao nível de *data governance*;
- **Processamento de língua natural:** aplicação de estudo da linguística para desenvolvimento de *software*;
- **Reasoner:** sistemas que permitem deduzir nova informação a partir de factos, como forma de implementação de raciocínio computacional;
- **Recommender system:** sistemas que permitem recomendar as melhores soluções para um utilizador a partir das suas preferências e perfil e face a um universo de alternativas recorrendo a modelos preditivos ou baseado em filtros sobre conteúdos.

Em adicional a estes sistemas, destaca-se igualmente formas de exploração de informação em *business intelligence* através de *reporting* e *dashboard*, quer sejam com estruturas de visualização fixas quer seja por estruturas *ad-hoc* no que é hoje designado por *self-service*.

3.4.4.2 *Ontology learning*

Pela sua forma uniforme de representação as ontologias têm uma capacidade de integração de diferentes domínios, enquanto construção de novas ontologias a partir de reutilização em outras ontologias, utilizando-se vários tipos de processos para o efeito como referido por Krummenacher e Strang (2007):

- **Ontology mapping:** expressão formal que descreve a relação semântica entre conceitos de diferentes ontologias;
- **Ontology alignment:** ligação semântica (mapeamento) entre ontologias que descrevem o mesmo domínio ou domínios relacionados;
- **Ontology merging:** construção de uma nova ontologia a partir de outras, criando uma ontologia mais geral sobre o domínio.

Apesar da existência de processos de criação de ontologias, linguagens e ferramentas, a sua criação é um processo complexo e moroso que exige conhecimentos avançados em determinados domínios sendo que o resultado pode ser inadequado, como referido por Hazman et al. (2011). Para colmatar este problema utilizam-se ferramentas para que os especialistas no domínio possam construir a ontologia, ou recorre-se a processos automáticos ou semiautomáticos para o efeito, no que é designado por *ontology learning*, segundo Hazman et al. (2011).

Hazman et al. (2011) definem *ontology learning* como um conjunto de técnicas e métodos para construir ontologias de raiz, enriquecendo ou adaptando uma ontologia base de forma semiautomática a partir de várias fontes. Os métodos utilizados correspondem a aprendizagem automática, aquisição de conhecimento, processamento de língua natural, descoberta de informação, inteligência artificial, raciocínio lógico e inferência, além de gestão de bases de dados.

De acordo com Hazman et al. (2011), utilizam-se vários métodos como *machine learning*, *knowledge acquisition*, *natural-language processing*, *information retrieval*, *artificial intelligence*, *reasoning* e *database management*, e por outro lado podem ser classificados a partir dos seus sistemas fonte para a criação de ontologias da seguinte forma:

- **Dados não estruturados:** livros, jornais e outros repositórios com dados em texto natural;
- **Dados semiestruturados:** dados com representação em linguagens como HTML e XML;
- **Dados estruturados:** dados representados em bases de dados e dicionários.

De acordo com Maedche e Staab (2001), o conceito de *ontology learning* integra várias disciplinas para permitir construir de forma automática ou semiautomática uma ontologia, incluindo mecanismos de aprendizagem automática. Para o efeito definiu um referencial onde se torna claro o ciclo de vida de atividades em torno da ontologia, sendo que no centro estão as ontologias de domínio com uma interface de *ontology learning* e uma arquitetura de solução, como apresentado na Figura 3.36 e Figura 3.37.

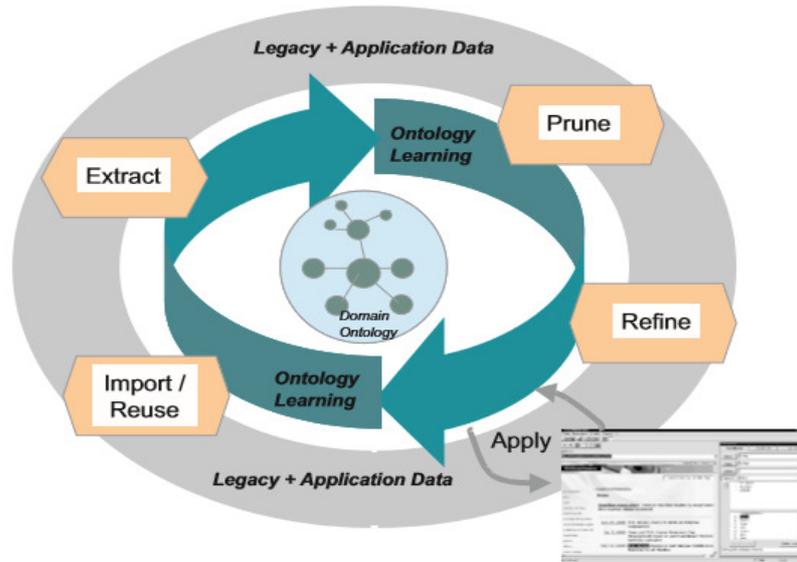


Figura 3.36: Processo de ontology learning (Maedche e Staab, 2001)

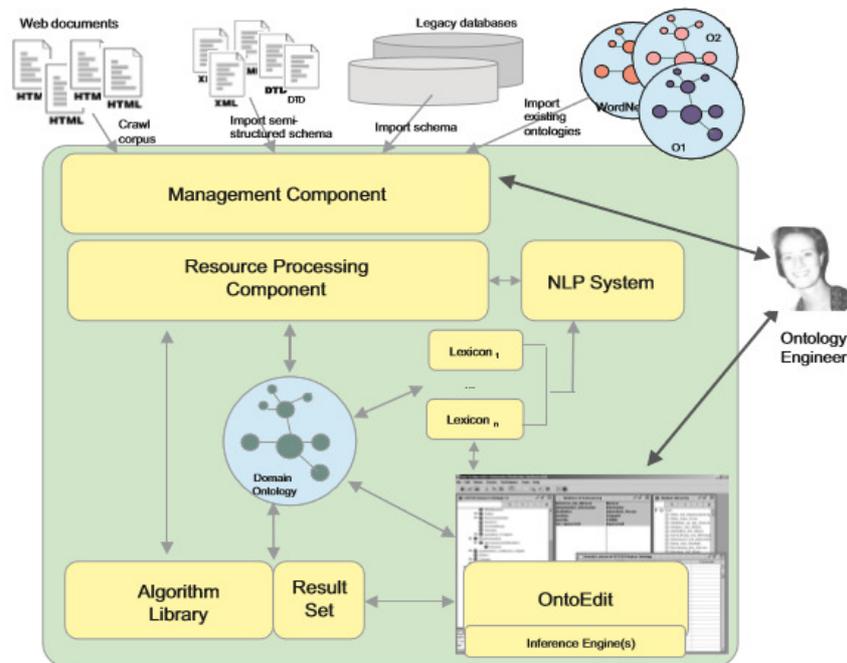


Figura 3.37: Componentes ontology learning (Maedche e Staab, 2001)

3.4.4.3 Data lineage

No contexto do *business intelligence* e *knowledge discovery*, torna-se necessário agregar fontes diversas de dados para consolidar informação para efeito de análise e tomada de decisão. Em particular no domínio do *business intelligence* um dos maiores desafios está relacionado com esta necessidade de integrar informação de fontes heterogéneas em formas de representação, conceitos e tecnologia segundo Inmon e Elmasri citados por Webjornsen (2005). Para se tratar os dados recorre-se a processos normalmente designados por *data integration*, que em *business intelligence* podem ser implementados via ETL (*Extraction, Transform and Load*) ou *data replication*, mas que se podem estender a transformações ao nível da preparação de dados para exploração até ao relatório ou *dashboard*.

As transformações resultantes dos processos de ETL ficam armazenadas em vários formatos de dados, tornando-se uma base de informação designada por metadados técnicos. A análise destas transformações de dados desde a origem até ao seu formato final recorre ao conceito de *data lineage* segundo Cui, Eckerson e Woodruff citados por Webjornsen (2005). Torna-se assim possível implementar processos de *tracing* dos dados, o que normalmente em soluções de *business intelligence* se torna crítico para responder a questões sobre a origem e transformação dos dados apresentados nos relatórios finais. No entanto, pela sua importância e modelo de funcionamento, torna-se crítico igualmente para sistemas operacionais como parte de abordagens transversais de *data governance*.

Ikeda e Widom (2009) abordam o problema de *data lineage*, face à complexidade de capturar regras de transformação em código computacional ou configurações diversas, além de que as transformações são conjugadas em vários processos, tornando necessário ligar as relações através de um grafo, como apresentado na Figura 3.38.

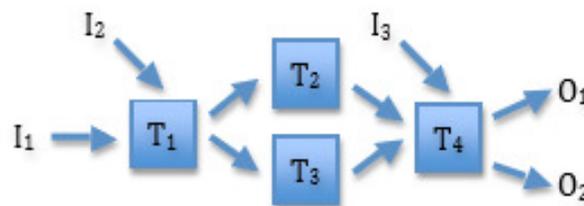


Figura 3.38: Gráfico de transformação (Ikeda e Widom, 2009)

Tendo por base a análise de várias alternativas, Ikeda e Widom (2009) sistematizam dois tipos de *data lineage* de acordo com tipo de respostas:

- *Where-lineage* ou *schema-level*, que tenciona responder a questões sobre que estruturas de dados são utilizadas nas transformações entre entrada-saída;
- *How-lineage* ou *instance-level*, que tenciona responder a questões sobre como determinadas instâncias de dados são transformados entre entrada-saída.

Simmhan et al. (2005) definem por sua vez uma taxonomia de características de *data lineage* aplicado a vários tipos de sistemas, focando nas razões pelos quais registam dados de *data lineage*, o que descrevem, como representam os dados e como permitem acesso a este tipo de dados, que designamos por metadados, tal como apresentado na Figura 3.39.

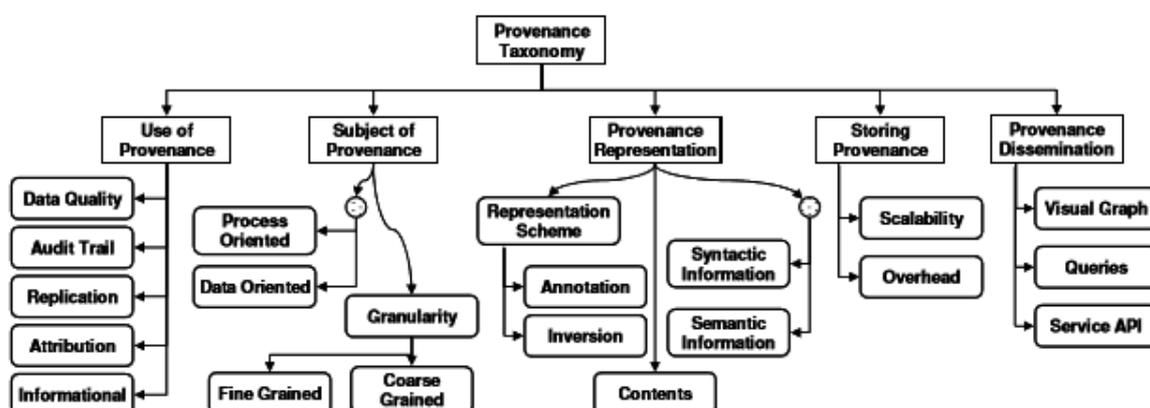


Figura 3.39: Taxonomia características *data lineage* (Simmhan et al., 2005)

Simghan et al. (2005) ao abordarem taxonomia, consideram que a representação dos metadados de *data lineage* depende da escalabilidade face ao potencial crescimento do volume deste tipo de dados, além de tornar claro que se deve limitar a dependência de intervenção manual na captura destes dados. Por outro lado, no que concerne à disseminação dos dados, foca essencialmente a utilização de grafos.

3.4.4.4 Processamento língua natural

Para Mote (2002), o processamento de língua natural tem por objetivo implementar em computação uma forma de entender e dialogar com utilizadores numa linguagem humana. Como tal, é uma das componentes da linguística computacional, com enfoque no tratamento computacional da linguagem e com aplicação prática em domínios como corretores ortográficos e gramaticais, além de tradutores e *parsers*. Othero (2006) cita no seu artigo, Vieira e Lima, para o qual a linguística computacional pode ser entendida como a “área de conhecimento que explora as relações entre linguística e informática, tornando possível a construção de sistemas com capacidade de reconhecer e produzir informação apresentada em linguagem natural”. Este tipo de linguística pode ser decomposto da seguinte forma:

- **Linguística de corpus:** centrada na utilização de coleções de textos, designados como “corpora” ou coleções, que normalmente são estruturados como bases de dados para representar uma amostra num determinado domínio e que podem ser focados em linguagem falada, linguagem escrita literária e textos de imprensa, entre outros. Estes “corpora” são organizados para possibilitar estudos de fenómenos linguísticos, como é o caso do *Wordnet* para a língua inglesa;
- **Processamento de língua natural:** desenvolvimento de *software*, como é o caso de *parsers* (analisador sintático), tradutores automáticos, reconhecedores de voz, geradores de resumos de texto ou *chatterbots* (interface para criar interação entre homem-máquina através de diálogo em língua natural).

Tendo por base a linguística, no processamento de língua natural deve-se considerar que na análise de um discurso, existem dois tipos de estudos:

- **Estrutura das palavras:** recorre-se à morfologia (estrutura interna das palavras), lexicologia (palavras que formam um idioma) e sintaxe (como são combinadas palavras para formar frases);
- **Estudo do discurso:** fonética (sons da linguagem), fonologia (padrões dos sons básicos de uma língua), semântica (sentido das frases e das palavras que integram essa frase), pragmática (forma como as oralizações são usados em atos de comunicação), estilística (estilo na linguagem) e terminologia (léxicos especializados das ciências e das técnicas).

Em particular, o estudo da lexicologia, sintaxe e semântica, assumem um papel crítico no contexto do processamento de língua natural. Neste sentido, Mote (2002) refere que existem vários módulos e métodos numa arquitetura de processamento de língua natural, que se podem resumir nos seguintes componentes:

- **Speech recognition:** conversão de uma frase numa sequência de palavras;
- **Language understanding:** análise de palavras no contexto de uma frase para identificar o seu significado;

- **Dialogue management:** interface de diálogo com o utilizador;
- **Communication with external system:** interface automática com outros sistemas, como é o caso de bases de dados, ontologias ou outros sistemas;
- **Response generation:** especificação da mensagem a ser gerada pelo sistema.

Tendo por base estes módulos, pode-se focar o *language understanding* como módulo central em investigação, que utiliza um processador, tal como referido na Figura 3.40, com as seguintes funções:

- **Lexical analysis:** examina as palavras para efeito do seu reconhecimento e morfologia, agrupando-as em *tokens*. Considerando cada palavra, pode-se considerar as palavras, que dependendo de utilização em frases assumem formas invariáveis (preposições, conjunções, advérbios), variáveis (artigos, pronomes e verbos) ou mistos (adjetivos e substantivos);
- **Syntax analyzer:** efetua o *parsing* e *post-tagging* das palavras recebidas do *lexical analysis*, agrupando-as em frases gramaticais. Nesta fase, ainda não se tem o significado, mas tem-se somente as palavras estruturadas como *templates*, isto é, em formas gramaticais corretas;
- **Semantic analyser:** aplica a gramática, enquanto conjunto de regras, para se obter expressões. Utiliza associação com frases humanas descritas em forma de código, utilizando conceitos de ontologias entre outras formas, para o efeito.

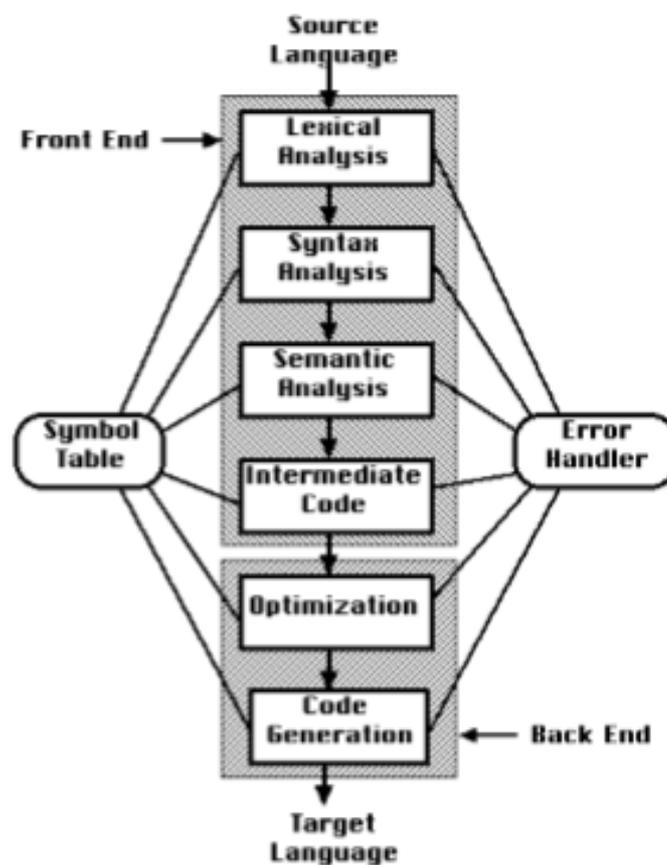


Figura 3.40: Processador NLP de Nikos Drakos citado por Mote (2002)

Devido à necessidade de se efetuar a análise semântica recorrendo a formas computacionais de comparação com frases humanas para entender o seu sentido, Bellandi et al. (2010) utilizam uma base de conhecimento estruturada como ontologia num sistema de processamento de língua natural para extração de relações a partir de padrões em textos. Esta relação com ontologias e face ao mecanismo de interpretação da semântica da linguagem, leva a uma aproximação entre os sistemas de processamento de língua natural e aprendizagem automática.

Para implementar sistemas de processamento de língua natural, existem várias ferramentas, como é o caso das seguintes:

- **NLTK:** plataforma construída em linguagem Python. Implementa interfaces com recursos lexicais como o WordNet, além de programas específicos para classificação, *tokenização*, *tagging*, *parsing* e análise semântica;
- **OpenNLP:** biblioteca Java para o processamento de linguagem natural em texto. Suporta tarefas como tokenização, segmentação de sentença, etiquetagem morfossintática, extração entidade nomeada, extração de sintagmas, análise sintática e resolução de correferência;
- **GATE:** *software* de código aberto para processamento e análise de texto. Permite criar fluxo de processamento de texto para realização de tarefas de extração de informação por meio de construção de taxonomia. Suporta tarefas como etiquetagem morfossintática, reconhecimento de entidades, tratamento de coreferência ou anáforas, anotação de texto com base em ontologia;
- **StanfordNLP:** *software* escrito em Java para processamento de linguagem natural que pode ser incorporado em outros programas, desenvolvido de acordo com a abordagem estatística para resolução de grandes problemas da linguística computacional. Processa textos nos idiomas inglês, francês, alemão, chinês e arábico;
- **LX-Center:** desenvolvido pela Universidade de Lisboa, fornece um conjunto de programas para processamento de textos escritos em português (segmentador de frases, segmentados lexemas, anotador categorial, *parser* e gramática computacional), além de serviços *online* para testes (e.g. lematizador verbal, conjugador verbal, reconhecedor nomes próprios, *parser*, etiquetador papéis semânticos);
- **WordNET:** base de dados lexical do inglês, em que substantivos, verbos, adjetivos e advérbios são agrupados em conjuntos de sinónimos, cada um expressando um conceito distinto. Os conjuntos estão interligados por meio de relações semântica-conceitual e léxica. Disponibiliza um *software* livre para navegação na base de dados;
- **UIMA:** biblioteca de *software* C++ e Java de código aberto para processamento de texto não-estruturado. Geralmente utilizada para criar sistemas de busca. Extensível através de *plugins*. Configurável por meio de arquivos XML para criação de fluxos de dados. Suporta extração de *tokens*, lematização, etiquetagem morfossintática, deteção de entidade e deteção de relações;
- **LingPipe:** biblioteca Java para processamento de texto usando linguística computacional. Suporta tarefas de reconhecimento de entidades (pessoas, organizações ou locais), classificação automática, correção ortográfica.

3.4.4.5 Reasoning

Segundo Bottou (2014), pode-se definir *reasoning* como “*algebraically manipulating previously acquired knowledge in order to answer a new question*”, o que envolve modelos de inferência *first-order* ou probabilístico. Bottou (2014) sugere igualmente abordagens para *machine-learning* mais próximas de ontologias de domínio reutilizadas para inferência ao referir que “*instead of trying to bridge the gap between machine learning systems and sophisticated “all-purpose” inference mechanisms, we can instead algebraically enrich the set of manipulations applicable to training systems and build reasoning capabilities from the ground up.*” *Reasoning* faz parte da lógica enquanto ciência que estuda o raciocínio no sentido de encontrar formas de pensar que conduzam a conhecimento verdadeiro. Normalmente a partir de um conjunto de factos conhecidos, designado por premissas, obtém-se uma conclusão correta. Estas premissas fazem parte do estudo de ontologias, nomeadamente na criação de *statement* sujeito-predicado-objeto que permitem registar conhecimento a partir do qual se gera novo conhecimento. Tal é o caso de “Francisco é Particular” e “Particular é Cliente”, então pode-se deduzir que “Francisco é Cliente” por transitividade.

Para se implementar sistema de inferência recorre-se a factos definidos numa ontologia onde se define as classes e relações para se formalizar um domínio, sobre o qual se aplica um conjunto de regras para inferir desta forma novo conhecimento. É o caso de motores de inferência como JESS, ApacheJENA, Pellet ou linguagens como SWRL.

3.4.4.6 Recommender system

Recommender system são tipos de sistemas que permitem aplicar funções designadas por *Utility System* sobre um universo de utilizadores e universo de itens para recomendar o item adequado a determinado utilizador ou lista de utilizadores. Como tal é um problema que se pode formalizar como *UtilityFunction* (*ListaUtilizadores*, *Listaltens*), sendo que o resultado é uma lista de itens adequado para a lista de utilizadores, tal como referido por Adomavicius e Tuzhilin (2005), que acrescentam que para implementação de sistemas deste tipo, recorre-se a três abordagens tipo, baseados normalmente em mecanismos de filtragem:

- **Collaborative:** utiliza uma abordagem baseada em *memory-based* para capturar a classificação de itens por utilizadores para chegar a uma conclusão, ou em alternativa recorre a *model-based* para utilizar modelos de aprendizagem preditivos para chegar à conclusão de recomendação;
- **Content-Based:** recorre a conteúdos de opinião dos utilizadores sobre os itens, que reflete a sua preferência e o conteúdo caracterizador dos itens, para cruzar as semelhanças e gerar uma recomendação;
- **Hybrid:** recorre a combinação de recomendações via *collaborative* e *content-based* para originar a recomendação final.

3.4.4.7 Data visualization

Telea (2007) refere que *data visualization* tal como *computer graphic application*, têm por objetivo representar dados de acordo com vários aspetos relacionados com esses mesmos dados. Como tal, para este autor, *data visualization* permite endereçar o problema da representação gráfica de dados.

Este domínio é igualmente visto como *visual analytics* por Keim et al. (2010), mas neste caso colocam o enfoque na combinação entre ferramentas analíticas e técnicas de visualização interativas para suporte à decisão, tendo por base grandes volumes de dados.

Data visualization ou *visual analytics* é utilizado em *business intelligence* no que concerne à exploração de dados ou na visualização de modelos, como é o caso da arquitetura empresarial, para personalizar e visualizar informação crítica da organização. Como tal, no contexto da nossa investigação, são essas as perspetivas de visualização que são críticas para análise.

Para sistemas de *business intelligence*, *data visualization* corresponde à forma como se pode explorar os dados, sob a forma de relatórios e *dashboard*, centrado na visualização de métricas agrupadas tematicamente e permitindo interatividade de visualização permitindo navegar pelos dados via dimensões. Esta navegação, depende de funções OLAP como é o caso de *slicing*, *dicing*, *rolling up* e *drill down*, como refere Van Der Aalst (2013), ao analisar estas funções para *data mining*, mas cujo modelo de funcionamento se aplica a visualização de dados em *dashbord*. O conceito de *dashboard* é igualmente consolidado por Elias (2012) para se centrar no problema de visualização de dados multidimensionais utilizando *dashboard*, considerando este modelo de visualização como representação visual de informação crítica para atingir objetivos, isto é, para suporte à decisão. Elias (2012) apresenta a estrutura de um *dashboard* com as suas várias componentes, tal como representado na Figura 3.41. Apesar de ser um *dashboard* com a tecnologia *SAP Business Objects*, representa um modelo geral de *dashboard* com a sua zona “A” com vários objetos gráficos que representam informação com várias perspetivas e permitindo navegação *drill* e *slice*, zona “B” com vários objetos e configurações passíveis de serem igualmente aplicados, zona “C” que mostra os filtros aplicados e a zona “D” com funcionalidades gerais de pesquisa, gravar, partilhar, entre outras



Figura 3.41: Estrutura de um *dashboard* (Elias, 2012)

Analisando *data visualization*, Frosch-Wilke e Tuchtenhagen (2016) consideram a visão de arquitetura empresarial como a solução para o alinhamento entre o negócio e os sistemas de informação, mas focando o problema de o sucesso da arquitetura empresarial estar dependente de soluções para visualizar de forma adequada e personalizada os modelos e conceitos de arquitetura empresarial. Para resolver o problema, propõe a utilização de um sistema de *business intelligence*, visto que este tipo de sistemas têm a função de disponibilizar informação de gestão centralizada e personalizada, integrando várias fontes de dados. Como tal, propõe um modelo de BI que importa dados de arquitetura empresarial para permitir a sua visualização. Para a sua implementação utilizam o sistema IBM *Cognos* e uma modelação de dados em *star schema* para representar os conceitos e visualizar sob a forma de *dashboard*. Esta abordagem difere, no entanto, da proposta na nossa investigação, pois não integra numa ontologia os metadados da arquitetura empresarial com bases de dados e glossários, para permitir reutilizar esta forma de ontologia para exploração numa lógica de ontologia empresarial.

Para relacionar *business intelligence* com arquitetura empresarial para efeito de *data visualization*, recorreremos de forma estrutural, a Buckl et al. (2010) que investigam formas de visualização *ad-hoc* de modelos, como é o caso de modelos de arquitetura empresarial, partindo do esquema apresentado na Figura 3.42.

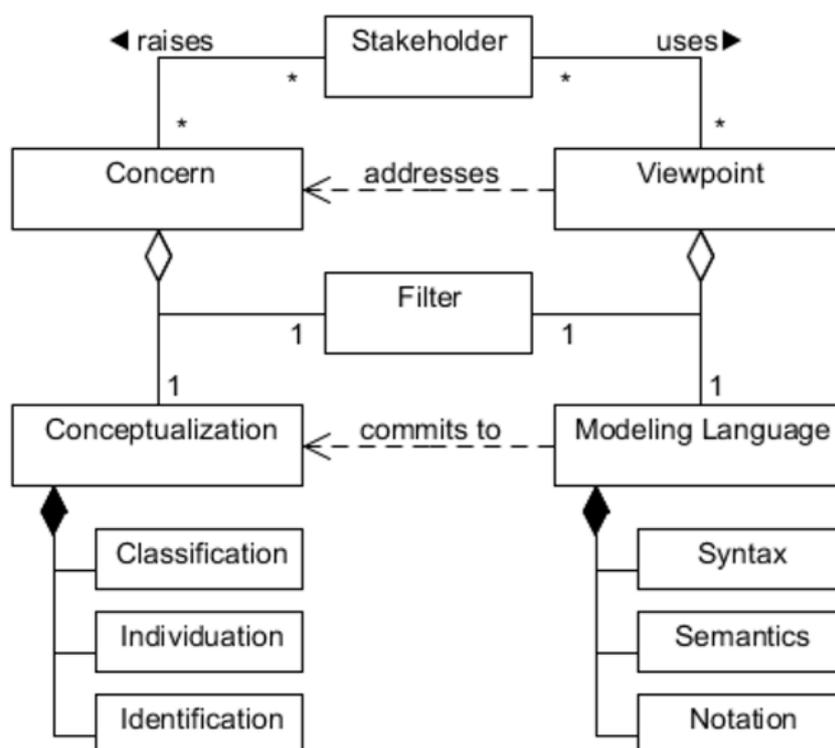


Figura 3.42: Pessoas, domínios de interesse e pontos de vista em *data visualization* (Buckl et al., 2010)

De acordo com a Figura 3.42, Buckl et al. (2010) referem que as pessoas (*stakeholder*) necessitam de entender domínios de conhecimento (*concern*) utilizando vários pontos de vista (*viewpoint*), aplicando filtros. Ao nível de domínios de conhecimento, recorre-se a modelos de ontologia (*conceptualization*) e ao nível de perspectivas recorre-se a linguagens de notação.

Para a nossa investigação, este modelo permite endereçar as vantagens de podermos capturar a arquitetura empresarial sob a forma de ontologia, permitindo a sua exploração, isto é, *data visualization* sob a forma de *dashboard* em *business intelligence*, de forma visual, personalizada e detalhada, mais adequada, face ao modelo de visualização gráfico de uma arquitetura empresarial, onde se utiliza representação de objetos (com várias formas gráficas) e linhas de relação entre os mesmos para representar a relação entre conceitos, mas que não permitem interação, personalização (aplicando filtros) e decomposição (como é o caso de funções OLAP como *drill* em *dashboard*).

3.5 Interoperabilidade de sistemas

A integração, é um componente crítico em engenharia organizacional, segundo Panneto e Molina (2008). Nham e Ishikawa (2005), acrescentam que esta criticidade tem a ver com o facto de que as organizações dependem da sua capacidade de comunicar, sendo que este processo envolve não só o processamento da informação, mas igualmente a capacidade de transferir a informação, no momento certo e para as pessoas certas. Para Whitman et al. (2006), uma das formas de garantir a integração é recorrendo ao conceito de interoperabilidade na medida em que os sistemas tenham capacidade de “interoperar” os seus dados, recursos e processos comerciais através de uma semântica definida para o contexto, independente dos formatos dos dados, interfaces, protocolos de comunicação e tipos de mensagens.

Para Whitman et al. (2006), as dificuldades da interoperabilidade prendem-se com questões técnicas (transporte de mensagens entre aplicações), sintática (formato de dados), semântica (interpretação de conteúdos das mensagens) e pragmática (vontade dos sistemas no trabalho colaborativo). Chen et al. (2008), resumem a dificuldade ao nível técnico que corresponde a questões técnicas e sintáticas indicadas por Whitman et al. (2006), nível semântico, e refere ainda o nível organizacional que corresponde à questão pragmática.

Estas dificuldades, segundo Whitman et al. (2006), prendem-se com o facto de as questões semânticas estarem fortemente relacionadas com os aspetos culturais e linguísticos, sendo desta forma barreiras à interoperabilidade, com particular destaque na interoperabilidade organizacional, onde é mantido o nível técnico, mas sendo reduzido apenas a um nível semântico. Na realidade, grande parte da informação das empresas reside em sistemas de informação e por outro lado as organizações têm uma dispersão geográfica interempresarial ou intraempresarial, com várias localizações e eventualmente com vários sistemas. Como tal, a arquitetura das empresas assume um papel crítico em processos de integração, segundo Chen et al. (2008), considerando o seguinte:

- As arquiteturas de sistemas permitem entender a lógica de sistemas e subsistemas em termos de estrutura e comportamento dos mesmos;
- A modelação das empresas baseado em referenciais, permite identificar conceitos em termos de estruturas, processos, sistemas e pessoas em tempo de desenho para facilitar a construção de sistemas com requisitos claros de integração física, entre aplicações e do negócio.

Esta capacidade nem sempre está presente nas organizações devido a questões relacionadas com formatos de representação de dados, quantidade e diversidade de fontes de dados e a própria história das organizações com vários sistemas adquiridos em diferentes fases da sua vida, tal como referido por Bucella citado por Ghawi (2010), agrupado nas seguintes dificuldades:

- **Autonomia:** considerando que os utilizadores podem aceder a partir de vários tipos de sistemas que podem estar desenhados sem considerar a necessidade de integração, podem comunicar com sistemas que não estavam previstos no início da sua implementação e podem ser executados de forma independente uns dos outros;
- **Distribuição:** os sistemas que processam os dados podem estar instalados fisicamente em vários tipos de ambientes de rede LAN ou WAN, sendo que o caso da internet leva a combinações infinitas de integração entre sistemas;
- **Heterogeneidade:** os sistemas têm modelos de dados diferentes com formas de representação distintas ou tecnologias de suporte distintas;
- **Instabilidade:** os sistemas têm ciclos de vida distintos que levam a que apareçam novas fontes de dados e formatos de representação que não são equivalentes entre os vários sistemas levando a instabilidade em integrações já existentes.

No entanto, a interoperabilidade, de acordo com Alexiev citado por Ghawi (2010) permite uma visão integrada de informação trazendo benefícios para o tratamento de informação de várias fontes ao nível de *data discovery* (localizar e relacionar a informação), *data integration* (processar informação) e *data quality* (verificação de consistência de dados).

Ghawi (2010) refere vários autores na compilação de abordagens para o problema de interoperabilidade sistematizando as seguintes abordagens:

- **Database-Translation:** integração através de mapeamento entre cada dois sistemas de acordo com os dados que tencionam integrar;
- **Standardization:** utilização do mesmo modelo ou norma de representação de dados e comunicação entre sistemas;
- **Federation e mediation:** federação de bases de dados para permitir que um sistema utilize vários sistemas de gestão de bases de dados como se fosse um único;
- **Language-based multi-base:** utilização de uma linguagem comum de manipulação de dados como o SQL para aceder aos dados;
- **Web-Services:** recurso a formas de API (*Application Program Interface*) para garantir um padrão comum de mensagens e serviços para integração entre sistemas.

Na tese, utilizamos a abordagem *standardization*, *database-translation* e *language-based multi-base*, na medida em que investigamos formas de reutilização de modelos entre dois tipos de sistemas, para garantir a interoperabilidade organizacional em torno de um modelo organizacional único construído como ontologia por *ontology learning*.

4 Caracterização do problema

Neste capítulo caracteriza-se o problema de interoperabilidade entre arquitetura empresarial e sistema de *business intelligence* relacionado com a ausência de uma visão holística baseada em metadados organizacionais, focando-se as causas e os impactos associados a este problema. Para o efeito, é apresentada a forma como se modela uma arquitetura empresarial (sub-capítulo 4.2), a forma como se implementa sistemas de *business intelligence* (sub-capítulo 4.3) tendo por base as ferramentas utilizadas na tese, para de seguida se sistematizar o problema (sub-capítulo 4.4), as causas e os impactos face à ausência de interoperabilidade (sub-capítulo 4.5).

4.1 Introdução

A implementação de sistemas de informação, no que concerne a aplicações informáticas, parte de um conjunto de requisitos definidos pelo negócio, que são enquadrados em processos devidamente integrados num modelo da organização.

Focando na relação entre modelação da organização e sistemas de *business intelligence*, segundo Gofarelli e Rizzi (2009), existem duas abordagens para implementação de sistemas de *business intelligence*. No caso de *top-down*, considera-se uma sequência de análise de requisitos, desenho e implementação, mas que tem custos elevados pois torna-se difícil obter todos os requisitos enquanto fontes de dados e regras de transformação. Por essa razão, indica que se utiliza mais a abordagem *bottom-up*, focada na criação de dados e respetiva exploração, tendo por base a identificação e análise específica de dados. No entanto, estes requisitos difíceis de obter na abordagem *top-down*, correspondem a conceitos enquadráveis em tipos de expressões de interrogação de negócio:

Todos os meses pretendo analisar a quantidade de vendas de produtos e quantidade de clientes agrupados por mês, canal comercial e segmento de cliente.

Esta expressão pode ser simplificada numa frase do tipo “analisar quantidade de produtos e quantidade de clientes, DE vendas POR mês, canal e segmento PARA mês atual”, com as seguintes particularidades:

- *Nesta frase simplificada, pode-se detetar um padrão onde primeiro se descreve as métricas “analisar quantidade de produtos e clientes”, seguida da palavra “DE” que separa a descrição de domínios “vendas”, seguida da palavra “POR” que separa a descrição de dimensões “mês, segmento e canal”, seguida da palavra “PARA” que separa as restrições “mês atual”. Existem palavras que não são relevantes (e.g. analisar, e, ou), sendo que as palavras restantes são conceitos, que correspondem a tabelas (caso de domínios) e campos em tabelas (restantes casos). Algumas palavras podem ter vários significados (e.g. produto pode ser definido igualmente como oferta comercial);*
- *Na expressão, aparecem os conceitos de vendas (que está associado a um conceito de contratos e movimentos), clientes, produto, canal comercial e segmento, normalmente definidos em arquitetura empresarial. De notar que na linguagem da organização, termos como canal, são automaticamente conhecidos como canal comercial. O mesmo acontece com segmento, que é um termo utilizado para classificar clientes;*

- Face à expressão em causa, os conceitos definidos correspondem a tabelas de factos/métricas, normalmente definida em *query* SQL DDL, como TVendas(DataVenda, CodVenda, CodCliente, CodProduto, CodCanal, outros campos) e tabelas de dimensões associadas, como TCliente (CodCliente, CodSegmento, outros campos), TProduto(CodProduto, DescProduto), TCanal(CodCanal, DescCanal) e TSegmento(CodSegmento, DescSegmento);
- Com base neste modelo de conceitos e base de dados deveria ser possível criar uma *query* SQL DML “SELECT <campos dimensões e campos métricas> FROM <Tabelas> WHERE <Condição> GROUP BY <campos de dimensões>” para responder à questão. Neste caso as dimensões são o mês, segmento e canal comercial, e as métricas são as contagens de produtos e clientes.

Considerando o exemplo da expressão exemplo acima indicada, a *query* DML seria a seguinte, acrescentando-se desagregação por produto para detalhar mais a informação:

```
SELECT TCanal.DescCanal, TSegmento.DescSegmento, TProduto.DescProduto,  
count(distinct TVendas.CodCliente), count(distinct TVendas.CodProduto)  
  
FROM TVendas  
  
LEFT OUTER JOIN TCliente on TCliente.CodCliente = TVendas.CodCliente  
  
LEFT OUTER JOIN TSegmento on TSegmento.CodSegmento = TCliente.CodSegmento  
  
LEFT OUTER JOIN TProduto on TProduto.CodProduto = TVendas.CodProduto  
  
LEFT OUTER JOIN TCanal on TCanal.CodCanal = TVendas.CodCanal  
  
WHERE Month(TVendas.DataVenda) = Month(today)  
  
Group by TCanal.DescCanal, TSegmento.DescSegmento, TProduto.DescProduto
```

Como tal, ao se considerar requisitos para *business intelligence*, na realidade está-se a encontrar soluções para responder a várias questões de negócio, que podem ser descritas por expressões que têm um padrão próprio. Cada expressão utiliza conceitos que estão definidos ao nível da arquitetura empresarial, ou deveriam estar, para garantir que a forma como monitorizamos o negócio tem por base os mesmos conceitos utilizados na definição do negócio de acordo com o modelo da organização. Por essa razão, Towers (2005) refere que a arquitetura empresarial é utilizada com o objetivo de suportar a implementação da estratégia, com base no posicionamento de todos os componentes da organização (leia-se, conceitos e relações utilizados na organização) que serve de referencial para permitir identificar onde estamos e para onde queremos ir. O facto de focar igualmente como objetivo a monitorização do processo de transformação, associado à análise de “para onde queremos ir”, posiciona a arquitetura empresarial na relação com *business intelligence* para apoiar em processos de mudança, o que torna esta relação entre os dois sistemas ainda mais relevante. Mas se considerarmos que a modelação da organização é efetuada numa arquitetura empresarial e a sua monitorização é efetuada em sistemas de *business intelligence* como criar interoperabilidade entre os dois sistemas ao nível de metadados utilizados por cada sistema? Para detalhar o problema, nos sub-capítulos seguintes apresentamos a forma como se efetua a modelação de arquitetura empresarial e a implementação de sistemas de *business intelligence*, para de seguida se sistematizar o problema face à questão de investigação.

4.2 Modelação de arquitetura empresarial

Tendo por base o modelo de Osterwalder e Pigneur (2010) e a sua adaptação para uma arquitetura empresarial tal como apresentado na Figura 3.15 do sub-capítulo 3.3.3.1, pode-se considerar que existem três áreas chave na arquitetura empresarial, tal como apresentado na Figura 4.1. A estratégia permite definir os objetivos, visão, missão e objetivos estratégicos. A orgânica permite definir os conceitos relacionados com custos associados ao que é necessário para manter a organização em funcionamento. O negócio corresponde aos conceitos relacionados com proveitos que permitem atingir os objetivos financeiros da organização. Na Figura 3.15 era referido igualmente a arquitetura funcional ou de processos, além da arquitetura de sistemas de informação. No entanto, para focar a nossa investigação, consideramos para análise somente estas três áreas chave pelo facto de serem as áreas de requisitos de negócio, enquanto que processos e sistemas de informação estão associados ao suporte a estes requisitos.

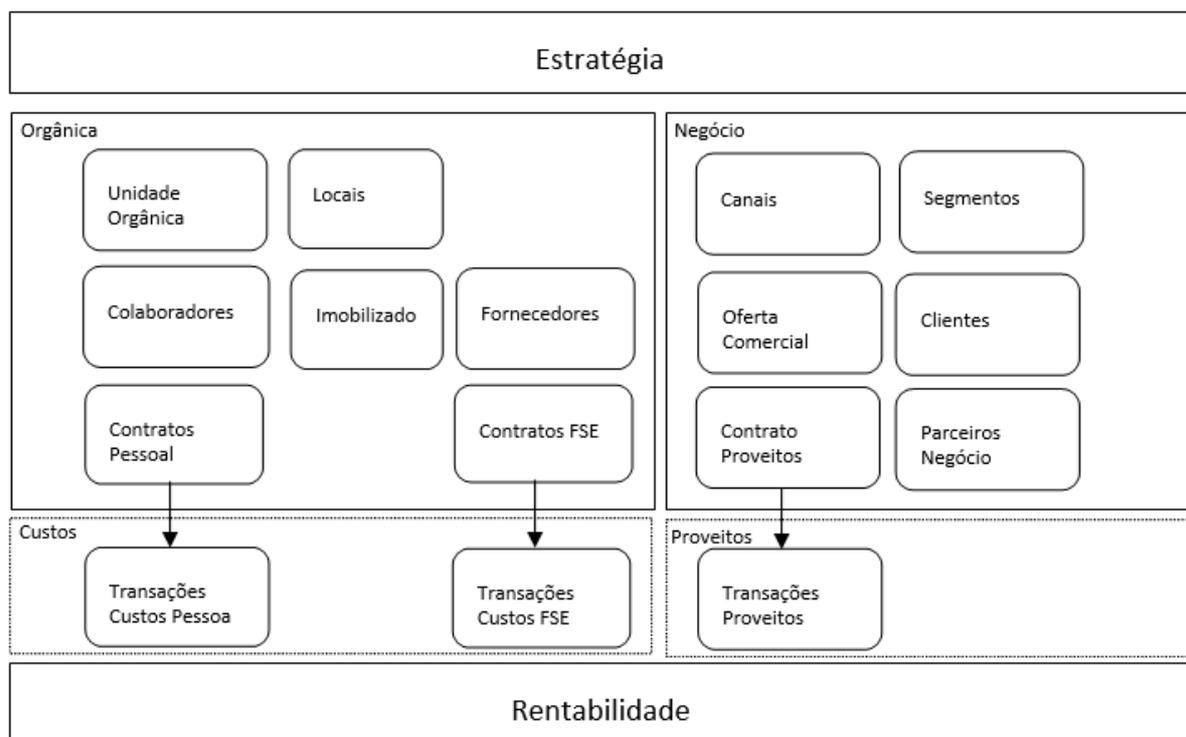


Figura 4.1: Adaptação do modelo de negócio de Osterwalder e Pigneur (2010)

Para se implementar este modelo, recorre-se a linguagens como ArchiMate e ferramentas como Archi (*ArchiMate tool*), que partem de princípios base de modelação considerando um metamodelo que define o tipo de conceitos, tipo de relações e que conceitos se podem relacionar com outros conceitos. Em ArchiMate existem os níveis de arquitetura *Business*, *Application* e *Technology* tal como detalhado no sub-capítulo 3.3.3.1. No caso de tipo de arquitetura *Business*, que corresponde ao modelo de negócio que consideramos detalhado entre arquitetura estratégica, arquitetura orgânica e arquitetura de negócio, existem elementos do tipo *active structure* (entidades que podem ter comportamento, como é o caso de um *actor* ou *location*), *behavior* (atividades realizadas por elementos *active structure*, como *process* ou *function*) e *passive structure* (elemento sobre o qual o *behavior* atua, como é caso de um *contract* ou *business object*).

Estes elementos, que passamos a designar por conceitos tem as seguintes características:

- Têm um conjunto de propriedades em pares *name-value*;
- Podem-se relacionar entre si recorrendo a tipos de relações (*structural relationship*) diversas, tais como *association*, *access*, *used by*, *realization*, *assignment*, *aggregation*, *composition*, *flow* e *triggering*;
- Podem ser visualizadas em modelos, considerados *views*, o que permite criar perspetivas, tais como organização, negócio, estratégia, custos e proveitos.

Esta forma de modelação e estrutura de visualização na ferramenta Archi é apresentada na Figura 4.2.

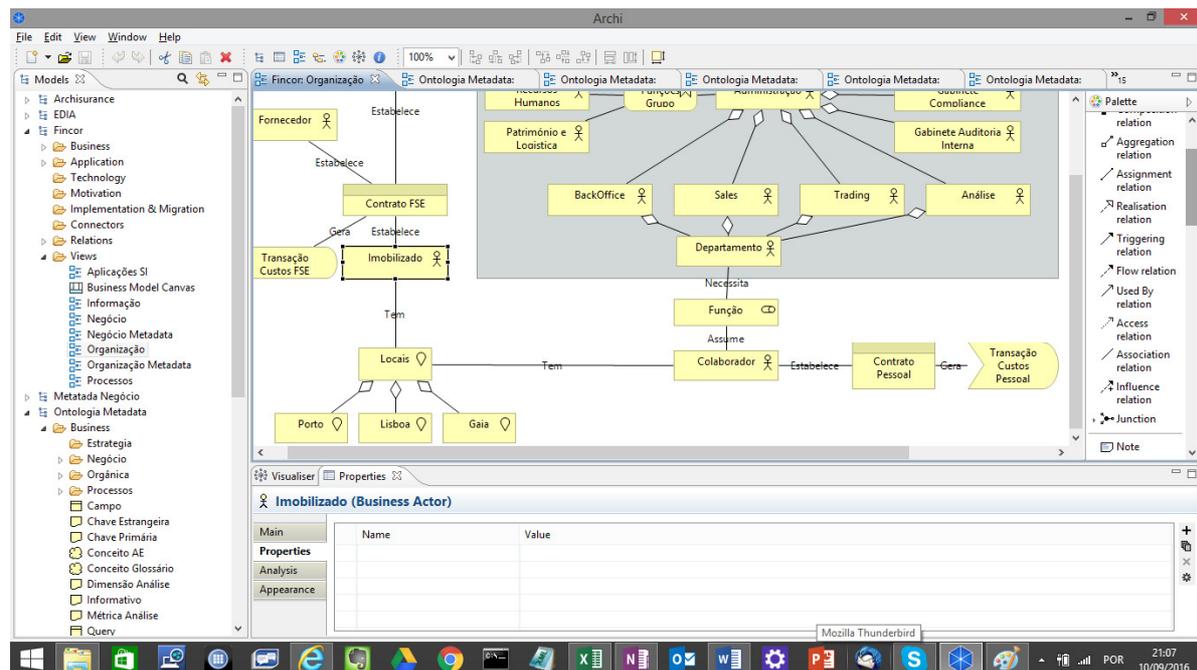


Figura 4.2: Modelação de arquitetura empresarial em Archi

No entanto, face à forma como é feita a modelação e a navegação, torna-se difícil resolver os seguintes problemas:

- Pode-se instanciar “locais”, pois são finitos, mas não se pode instanciar “clientes” ou “produtos”, mas somente a sua tipologia, pois senão, a visualização torna-se de difícil leitura e atualização;
- Pode-se detalhar a semântica da informação, indicando que os “locais” têm uma relação com “imobilizado” e com “colaboradores”, considerando que determinam que são espaços onde estes outros elementos existem. Mas não permite detalhar conceitos equivalentes ou tabelas associadas de dados relacionadas, sem se tornar o modelo complexo. No caso do modelo de dados, ter-se-ia que criar as tabelas como parte da arquitetura de informação e ligar cada conceito de negócio com as tabelas, vistas como um conceito de dados, criando depois uma *view* específica para o efeito. O mesmo aconteceria com os conceitos equivalentes, que neste caso seria uma arquitetura específica de glossário que normalmente não é representada em arquitetura empresarial, mas que teria a mesma dificuldade de visualização.

Apesar dos problemas acima indicados, ao se modelar, vai-se mantendo uma estrutura de dados na ferramenta, que permite exportar para ficheiros de texto (formato CSV) ou OWL. Assim, este modelo pode ser visto não como a totalidade dos metadados organizacionais, mas somente como uma parte dos mesmos, pois é necessário complementar com a seguinte informação:

- Glossários para determinar termos equivalentes a cada elemento modelado;
- Bases de dados para se entender a correspondência entre um conceito como “Imobilizado” e tabela ou campos onde é implementado.

Como tal, é necessário considerar que na arquitetura empresarial se captura os principais conceitos e suas relações como um metamodelo da empresa, visto como ontologia e sem instanciação, mas que pode servir para reutilização em outros sistemas.

4.3 Implementação de sistema business intelligence

Ao se implementar um sistema de *business intelligence* tem-se de considerar o seguinte:

- Que dimensões e métricas são relevantes?
- Que relações hierárquicas existem entre dimensões (e.g. distrito, concelho e freguesia)?
- Que fórmulas são utilizadas para cálculo de algumas métricas (e.g. Total Crédito = Capital + Juros + Impostos + Despesas)?
- Como organizar os processos de ETL?
- Como organizar os dados, considerando que é necessário identificar as tabelas e campos, modelos (e.g. relacional, *star schema*, *snowflake*) e níveis de dados (e.g. data staging, ODS, DW, DM)?
- Como organizar a camada de exploração de dados?

Normalmente, não se recorre à organização da informação derivada da arquitetura empresarial, pelo que depende de quem implementa ou define a arquitetura da solução. Este facto leva a que seja mais tarde difícil de conciliar uma visão entre o modelo da organização e a forma como se organiza a informação de gestão.

A implementação envolve processos de ETL e modelos de dados, além da exploração de dados. Focando na camada de exploração, pode-se notar as seguintes características, tal como apresentado na Figura 4.3, onde é apresentado um *dashboard* em Microsoft Power BI Desktop:

- A camada de apresentação é estruturada por tabuladores ou áreas de análise, que neste caso correspondem a “sheets” que deveriam ser equivalentes a uma perspetiva de análise de informação de gestão de acordo com a arquitetura empresarial;
- Normalmente, existe um painel para definir filtros genéricos para toda a perspetiva de visualização. Toda a visualização é orientada para combinação entre dimensões e métricas que estão expostas ou que se podem ir colocando por opção do utilizador;
- De notar que a própria visualização apresenta títulos por defeito como QtdCompras por Categoria ou QtdCompras por CodPais. Isto é “analisar <métrica> por <dimensão>”;

- No painel de visualização, pode-se criar várias análises correlacionadas entre si. Isto é, ao se selecionar um item numa das visualizações está-se a filtrar itens em outras visualizações. Por outro lado, uma das funcionalidades principais destas ferramentas é o designado por *drill*, que deriva do entendimento de modelos de dados multidimensionais conhecido como OLAP. Isto é, a possibilidade de se analisar a partir de um indicador, a sua decomposição (e.g. a partir do total de vendas, desagregar primeiro por país, depois por distrito e depois por concelho). Esta forma de visualização permite ir detalhando e navegando pelos dados de forma mais adequada e enriquecida com métricas, face a um modelo de instanciação de dimensões derivado de conceitos que são igualmente modelados na arquitetura empresarial.

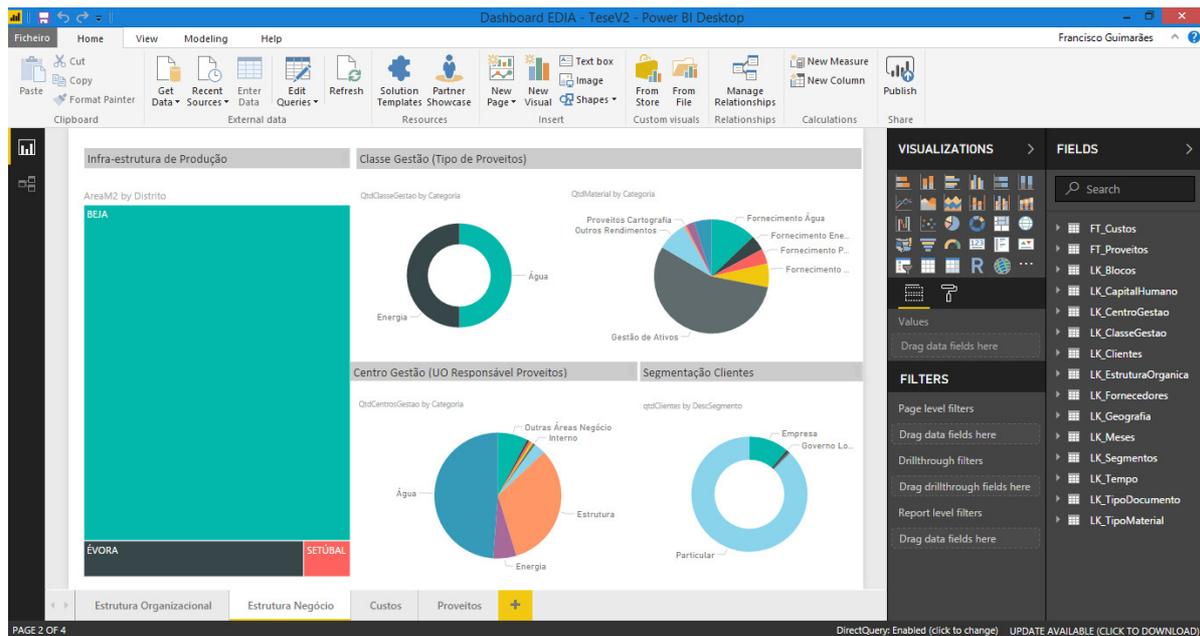


Figura 4.3: Dashboard informação de gestão em Microsoft PowerBI Desktop

4.4 Sistematização do problema

Os conceitos utilizados em sistemas de *business intelligence*, correspondem a métricas e dimensões, organizados em processos de carregamento de dados (ETL), armazenamento de dados (*data staging*, ODS, DW e DM) e exploração de dados (*dashboard* ou relatórios organizados de forma temática).

As áreas temáticas e conceitos, nomeadamente as dimensões, correspondem a elementos normalmente definidos durante a modelação de uma arquitetura empresarial. Isto porque na arquitetura empresarial é necessário identificar todos os elementos chave da organização e como se relacionam. Sendo a arquitetura empresarial uma forma de definição e acompanhamento da estratégia, necessita de indicadores de desempenho organizacional obtidos via *business intelligence*, mas com base nos mesmos termos e conceitos. Por outro lado, a própria visualização de detalhe da forma como a organização está modelada, não permite que se consiga visualizar numa arquitetura empresarial as instâncias de conceitos como clientes, contratos, produtos, mas somente a sua tipologia e regras, ou semântica de relações entre conceitos.

Atualmente, o alinhamento entre os dois tipos de sistemas não existe, quer em termos de metodológicos, quer em termos técnicos. Ao nível de ferramentas *open source* ou comerciais, não se identificaram tendências de inclusão de modelação de arquitetura empresarial em ferramentas de *business intelligence*, ou vice-versa, nem no contexto atual de suporte para *data governance* ou numa visão de metadados corporativos. O que existe, são tentativas de classificar as dimensões e métricas a partir do formato de dados nas ferramentas de *business intelligence* e tentativas de criar modelos complexos de arquitetura empresarial com detalhe de instâncias de conceitos (isto é, detalhar não só a relação entre unidade orgânica ao nível de direção, departamento, mas também as direções e departamentos concretos, sendo mais complexo no caso de conceitos de clientes e lista de clientes).

A nossa investigação tenta encontrar uma solução, utilizando a arquitetura empresarial como forma de criar uma ontologia base, sem detalhe de conceitos em termos de instâncias, enriquecida com outros conceitos de glossários e bases de dados, para ser reutilizada em sistemas de *business intelligence* enquanto metadados, utilizando ontologias OWL.

4.5 Causa e impacto

As causas do problema têm origem no seguinte:

- Ciclos de vida diferentes para a implementação de vários sistemas;
- Heterogeneidade de modelos e tecnologia de suporte aos vários sistemas;
- Ausência de normas para interoperabilidade de metamodelos das organizações;
- Ausência de definição de arquitetura de informação como um dos níveis da arquitetura empresarial, o que não permite uma ligação direta entre conceitos de negócio (metadados de negócio) e conceitos técnicos de dados (metadados técnicos);
- Ausência de processos e políticas para *data governance* focada na visão de metadados organizacionais e sua relação com arquitetura de informação;
- Não utilização dos modelos da organização instanciados na arquitetura empresarial para serem reutilizados na definição de dimensões e métricas, quer para os metadados técnicos, quer para os metadados de negócio em sistemas *business intelligence*.

O impacto deste problema reflete-se no seguinte:

- Desalinhamento entre a modelação da organização e avaliação de desempenho, não garantindo que o que é definido na arquitetura empresarial é depois utilizado em sistemas *business intelligence*;
- Desalinhamento de comunicação entre a estratégia definida e a forma como a representação evidencia a concretização desta estratégia ao nível de termos e conceitos utilizados nos sistemas de informação na interação com utilizadores;
- Dificuldade de manter sistemas de arquitetura empresarial face a alterações de conceitos implementados no sistema de informação e na obtenção de indicadores de gestão para serem visualizados em conjunto com o modelo da organização.

5 Questionário a empresas

Neste capítulo apresentam-se o resultado do questionário a nove organizações sobre a utilização de arquitetura empresarial (sub-capítulo 5.2), utilização de sistemas de *business intelligence* (sub-capítulo 5.3) e aferição sobre o problema e hipótese de solução em investigação (sub-capítulo 5.4).

5.1 Introdução

O questionário a empresas tinha como objetivo aferir sobre a importância do problema e adequação da hipótese de solução. Como tal, o questionário teve por base a necessidade de identificar os níveis de utilização de arquitetura empresarial e de *business intelligence*, além de um teste à hipótese de solução em torno da relação entre os mesmos. Para facilitar o entendimento e preenchimento do questionário, foi utilizado o *Google Forms* para preparar o questionário e enviado por *email* para as entidades em causa, tendo-se criado igualmente um *website eaintelligence*² como complemento de informação sobre a tese. Todas as questões foram estruturadas de acordo com o problema e hipótese de solução em investigação, tendo-se deixado sempre uma opção de resposta em aberto para identificar outras alternativas de resposta que os inquiridos pudessem indicar. Para o efeito, o questionário foi estruturado em três partes:

- Nível de utilização e complexidade de gestão da arquitetura empresarial;
- Nível de utilização e complexidade de gestão de sistemas de *business intelligence*;
- Perceção sobre o problema e hipótese de solução centrado na relação entre arquitetura empresarial e sistemas de *business intelligence*.

Não foi utilizada uma técnica de amostragem pois foram recebidas respostas de todas as nove entidades convidadas entre janeiro e maio de 2016 (quatro do sector financeiro, duas seguradoras, uma empresa de restauração internacional, uma empresa internacional de serviços móveis e uma empresa de telecomunicações). Na análise das respostas, apresentada nos sub-capítulos seguintes, nos casos em que os resultados são apresentados sob a forma de tabela (opções múltiplas) é sempre apresentada a quantidade de respostas e a respetiva percentagem face ao total.

5.2 Utilização de arquitetura empresarial

Na Figura 5.1 e Figura 5.2 são apresentados os resultados relativos à utilização e responsabilidade de gestão da arquitetura empresarial. A utilização a mais de 5 anos é indicada por 66% dos inquiridos, sendo que a sua gestão é assegurada pela área de sistemas de informação (90%) e pela área de organização e métodos (60%). De notar que a unidade de sistemas de informação é responsável pela arquitetura aplicacional, informacional e tecnológica, enquanto a unidade de organização e métodos é responsável pela arquitetura de processos e organizacional. Somente 40% dos inquiridos referem o papel da unidade de planeamento estratégico, que é normalmente responsável pela arquitetura estratégica e de negócio.

² <http://EAIntelligence.wordpress.com>

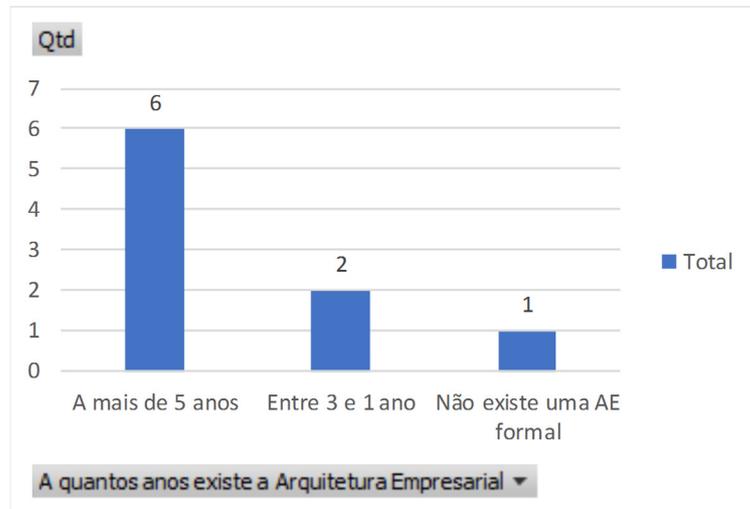


Figura 5.1: A quantos anos existe a arquitetura empresarial?

Que Unidades Orgânicas estão envolvidas na gestão da AE?

Planeamento Estratégico	4	40%
Organização e Métodos	6	60%
Sistemas de Informação	9	90%
Other	2	20%

Figura 5.2: Que unidades orgânicas estão envolvidas na gestão da arquitetura empresarial?

Nas Figuras 5.3, 5.4 e 5.5, são apresentados os resultados relativos a metamodelos utilizados e níveis de arquitetura. Neste caso, 50% dos inquiridos utilizam modelos proprietários das ferramentas e somente 20% utilizam o ArchiMate. Por outro lado, a maioria dos inquiridos refere que os níveis de arquitetura mais modelados são os sistemas de informação (média de 60% considerando os níveis aplicativos, informacionais e tecnológicos), seguido dos processos (60%) e por fim o modelo organizacional (40%). De notar a fraca utilização de modelos de negócio sem *balance scorecard* (10%) o que comprova a falta de alinhamento entre a arquitetura empresarial e sistemas de *business intelligence* pois sem este nível de arquitetura os conceitos base de negócio não estão alinhados com o que é implementado para monitorização do desempenho. Este facto é comprovado pela fraca utilização em fase de planeamento estratégico (10%) e mesmo para planeamento estratégico de sistemas de informação (40%) face à maior utilização para documentação de processos (60%). De notar ainda a utilização para análise de risco em sistemas de informação (40%) e na análise de impacto de alterações em sistemas de informação (50%), o que pressupõe a utilização destes sistemas como um catálogo de componentes e relações para se conseguir atingir este objetivo de análise de impacto em alterações ou em componentes de risco.

Que Metamodelo é utilizado para AE?

ArchiMate	2	20%
Zachman	1	10%
Proprietário da ferramenta	5	50%
Other	3	30%

Figura 5.3: Que metamodelo é utilizado para arquitetura empresarial?

Que níveis de arquitetura de AE existem no seu modelo?

Negócio sem BSC	1	10%
Negócio com BSC	0	0%
Orgânica	4	40%
Processos ou Funcional	6	60%
Aplicacional	7	70%
Informacional	5	50%
Tecnológica	6	60%
Other	1	10%

Figura 5.4: Que níveis de arquitetura de AE existem no seu modelo?

Qual a utilização da AE na Organização?

Em fase de planeamento estratégico de negócio	1	10%
Em fase de planeamento estratégico de SI	4	40%
Na análise de impacto de alterações em SI	5	50%
Na documentação de processos de negócio	6	60%
Na análise de risco de sistemas de informação	4	40%
Other	2	20%

Figura 5.5: Qual a utilização da arquitetura empresarial na organização?

A indicação de ferramentas considera somente os sistemas ARIS, Hermes/TIMM e Oracle, com grande destaque para não utilização de ferramentas (não aplicável ou nenhuma), tal como apresentado na Figura 5.6. Na realidade, em casos onde a resposta é “não é aplicável”, é colocado um comentário de que *“existe uma documentação dos processos, geridos na ferramenta QPR; Existe um repositório de aplicações, gerido no Changepoint; Existe uma documentação da DWH e do sistema informacional; Camada de integração, gerida no DataPower da IBM. Toda esta informação é mantida e acessível para referência.”*. Como tal, na realidade, em casos em que não é indicada uma ferramenta que centralize a arquitetura, é porque existem ferramentas diversas, para cada nível de arquitetura.

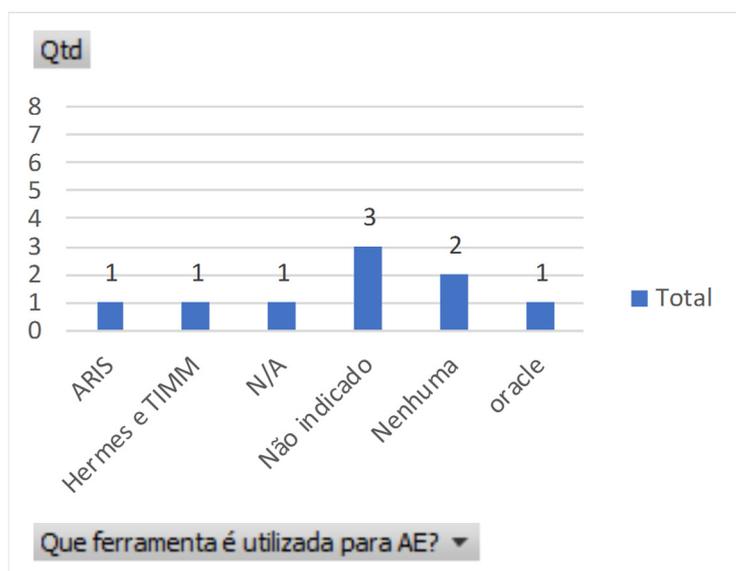


Figura 5.6: Que ferramenta é utilizada para AE?

No que concerne à integração da arquitetura empresarial com outros sistemas, a Figura 5.7 torna claro que tal ocorre somente em 50% dos casos, mas é de ressaltar as tentativas de integração com sistemas que têm metadados próprios, como os seguintes:

- Existem tentativas de integrar com sistemas que têm metadados técnicos próprios. Tal é o caso de sistemas SOA (40%), metadados de *business intelligence* (30%), sistemas de BPM/workflow (30%) e CMDB (30%);
- Existem tentativas de integrar com sistemas que têm metadados de negócio próprios. Tal é o caso de catálogo de produtos/serviços (30%) normalmente mais associado a arquitetura de negócio, ou o caso de capital humano (20%) mais associado a arquitetura organizacional.

A ferramenta tem integração com outras fontes de artefactos?

Sistemas de BPM/Workflow	3	30%
CMDB (Configuration Management DB)	3	30%
SOA	4	40%
Metadados BI	3	30%
Gestão de Capital Humano	2	20%
Catálogo de Produtos/Serviços	3	30%
Other	3	30%

Figura 5.7: A ferramenta tem integração com outras fontes de artefactos?

Nas Figuras 5.8, 5.9 e 5.10, são apresentados os resultados relativos às questões relacionadas com a gestão da arquitetura empresarial. É considerado que a criação da arquitetura empresarial tem um nível elevado de dificuldade (70%) além de um nível elevado de dificuldade de atualização (55.6%). Esta dificuldade está na sua maioria associada à dificuldade de envolver a organização (80%) e pelas alterações constantes de negócio (60%). Esta realidade demonstra a pertinência do problema associado a questões de alinhamento em torno dos mesmos conceitos entre modelação da organização (para criar mais envolvimento) e monitorização (para melhor captar as alterações).

Qual o nível de dificuldade na criação da AE?

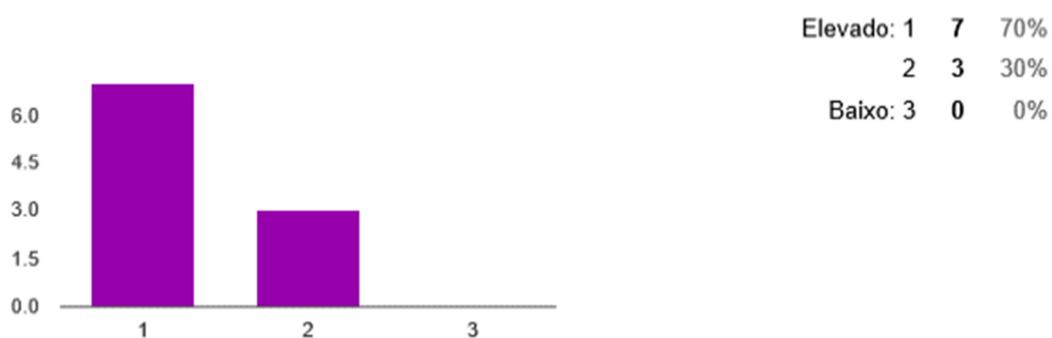


Figura 5.8: Qual o nível de dificuldade na criação da arquitetura empresarial?

Qual o nível de dificuldade na atualização da AE?

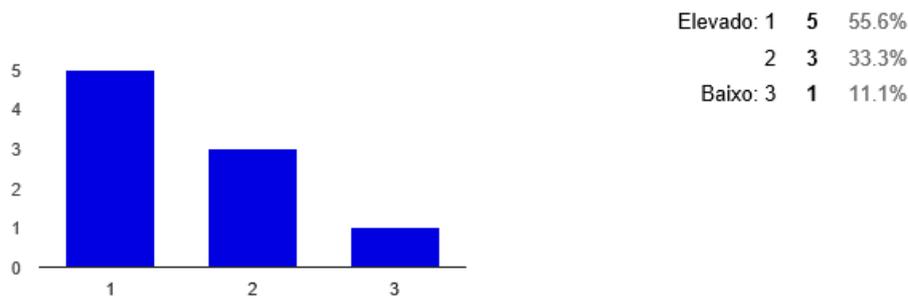


Figura 5.9: Qual o nível de dificuldade na atualização da arquitetura empresarial?

Qual o principal motivo na dificuldade de gestão da AE?

Envolvimento da organização	8	80%
Obtenção de informação de ativos da organização	3	30%
Obtenção de informação de gestão para KPI's	3	30%
Alterações constantes de negócio	6	60%
Other	1	10%

Figura 5.10: Qual o principal motivo na dificuldade de gestão da arquitetura empresarial?

5.3 Utilização de business intelligence

Nas Figuras 5.11 e 5.12, são apresentados os resultados relativos à utilização e responsabilidade de gestão de sistemas de *business intelligence*. Destaque-se a utilização a mais de 5 anos (36%) e “entre 1 e 3 anos” (33%). Esta arquitetura é na sua maioria gerida pela unidade de sistemas de informação (50%) ou unidades específicas criadas para o efeito (50%) para serem mais transversais na organização.

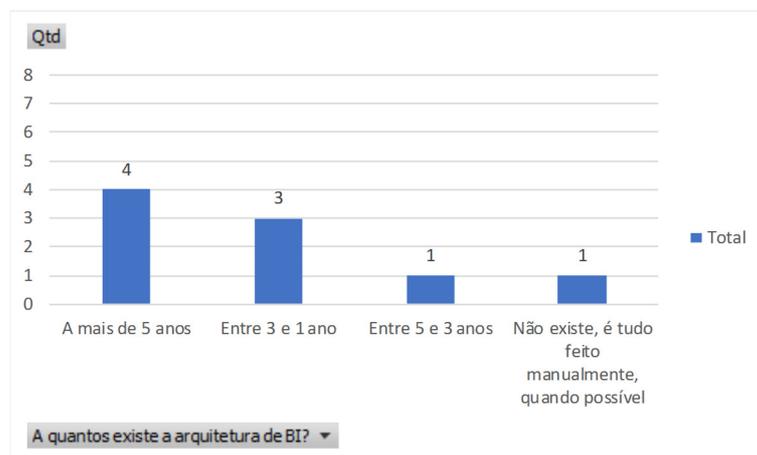


Figura 5.11: A quantos anos existe a arquitetura de *business intelligence*?

Que Unidades Orgânicas gerem a arquitetura de BI?

Unidade específica de BI	5	50%
Unidade de gestão de aplicações de SI	5	50%
Marketing	2	20%
Planeamento Estratégico	2	20%
Other	1	10%

Figura 5.12: Que unidades orgânicas gerem a arquitetura de *business intelligence*?

Como ferramentas de *business intelligence*, são utilizadas diversas, tal como a seguir apresentadas na Figura 5.13.

Que ferramentas são utilizadas para BI (ETL, Metadata e Exploração BI)?
SAS, PENTAHO, MicroStrategy
SAS, CA Repository, Oracle BI - (não exaustivo)
nenhuma
SAS
programas próprios
PL/SQL, Spago BI, Microsoft SQL Server, Microsoft Intergration Services, Microsoft Analysis Services, Microsoft Reporting Services
SSIS e SSAS + Excel + Sharepoint+ SSRS
Micorsoft SQL Server, Teradata e Microstrategy
sas

Figura 5.13: Que ferramentas são utilizadas para BI?

Nas Figuras 5.14, 5.15 e 5.16, são apresentados os resultados relativos a metamodelos utilizados e níveis de arquitetura. A maioria dos inquiridos (66.7%) indica que utiliza modelos desenhados pela organização. Ao nível de modelo de dados são utilizados DW (70%) mas também *data mart* (60%) com pouca utilização de ODS (20%). Ao nível de exploração de dados, a maioria utiliza *reporting* fixo (70%) seguido de *dashboard* (60%). De notar que nenhum dos inquiridos utiliza atualmente *Big Data*. Ao nível de metadados, deteta-se uma fraca utilização sendo utilizado essencialmente o que é gerado pelas ferramentas (40% para ETL e 30% para exploração de BI), devendo-se destacar utilização de metadados corporativos transversais entre ETL e exploração (30%) mas que não são transversais à organização (0%) onde se incluiria igualmente a arquitetura empresarial e sistemas operacionais. Este facto comprova a problemática em investigação nesta tese.

Que (meta)modelo é utilizado para BI?

Desenhado pela organização	6	66.7%
Adaptação de modelo do fabricante da ferramenta	2	22.2%
Other	1	11.1%

Figura 5.14: Que metamodelo é utilizado para BI?

Que níveis de arquitetura de BI existem?

ODS	2	20%
DataWarehouse	7	70%
DataMart	6	60%
Reporting Fixo	7	70%
Reporting em Self-Service	4	40%
Dashboard	6	60%
BigData Analytics	0	0%
Data Mining	3	30%
Function Analytics (Profitability, Customer Intelligence, Risk Intelligence, outros)	1	10%
Other	1	10%

Figura 5.15: Que níveis de arquitetura de BI existem?

Que tipo de metadata existe?

Não existe	3	30%
Gerada e utilizada no ETL	4	40%
Gerada e utilizada pelas ferramentas de exploração de BI	3	30%
Metadata transversal para ETL e Exploração BI	3	30%
Metadata Corporativa para BI e sistemas operacionais	0	0%
Other	4	40%

Figura 5.16: Que tipo de metadados existem?

Nas Figuras 5.17 e 5.18, são apresentados os resultados relativos à gestão de sistemas de *business intelligence*. A maioria (50%) considera um nível elevado de dificuldade de gestão, sendo que o principal motivo tem a ver com a dificuldade na captura de novas dimensões e métricas a partir dos sistemas operacionais (50%), constantes alterações nos sistemas operacionais sem acompanhamento pelas equipas de *business intelligence* (50%) e dificuldade na manutenção de metadados consistentes com conceitos de negócio (50%). Este facto comprova a adequação da problemática endereçada nesta tese.

Qual o nível de dificuldade na gestão de BI?

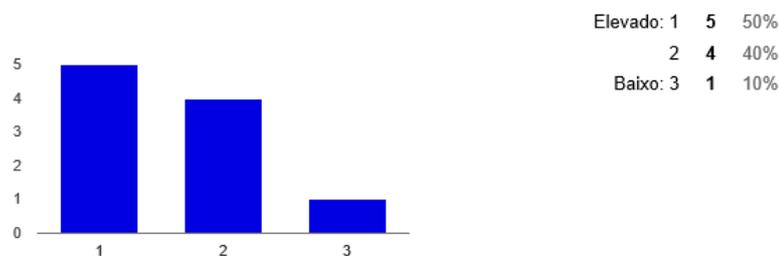


Figura 5.17: Qual o nível de dificuldade na gestão de *business intelligence*?

Qual o principal motivo de dificuldade de gestão de BI?

Velocidade de necessidades de utilizadores para novos Dashboard	3	30%
Velocidade de necessidades de utilizadores para novos Reporte's	3	30%
Captura de novas métricas e dimensões a partir de sistemas operacionais	5	50%
Alterações de sistemas operacionais sem acompanhamento pelo BI	5	50%
Manutenção de metadados consistentes entre conceitos de negócio e conceitos de BI	5	50%
Other	1	10%

Figura 5.18: Qual o principal motivo de dificuldade de gestão de *business intelligence*?

5.4 Perceção do problema e adequação da hipótese de solução

Na Figura 5.19 é apresentado o resultado relativo à importância de utilização do modelo de arquitetura empresarial como ontologia corporativa. É reconhecida a sua importância por conter todos os conceitos e relações do negócio (50%), mas complementada com glossários (50%) e modelos de dados (50%).

Qual o nível de importância do modelo AE como ontologia corporativa?

Não é relevante	0	0%
Relevante porque contém todos os conceitos e relações (semântica) do negócio	5	50%
Relevante mas deve ser complementado com glossários de termos	5	50%
Relevante mas deve ser complementado com modelos de dados (tabelas, campos e relações)	5	50%
Other	1	10%

Figura 5.19: Qual o nível de importância do modelo E como ontologia corporativa?

Nas Figuras 5.20 e 5.21, são apresentados os resultados relativos à relação entre conceitos da arquitetura empresarial, e dimensões e métricas em *business intelligence*. De destacar que 60% dos inquiridos consideram que facilitaria a gestão corporativa de metadados entre arquitetura empresarial e *business intelligence* e 80% considera a sua importância para a relação entre arquitetura empresarial e *business intelligence* no que concerne a métricas (KPI). É indicada ainda a importância de otimizar o esforço de modelação de dimensões e hierarquias em BI (40%). Estas conclusões demonstram a relevância na relação entre os dois sistemas enquanto problemática endereçada nesta tese.

Qual o nível de importância de classificar e reutilizar dimensões em BI a partir de conceitos e relações em AE?

	Não é relevante	0	0%
Não é possível porque em AE não existem todas os conceitos e relações correspondentes as dimensões em BI		2	20%
	Otimizaria o esforço de modelação de dimensões e hierarquias em BI	4	40%
	Facilitaria a gestão corporativa de metadados entre AE e BI	6	60%
	Other	2	20%

Figura 5.20: Qual o nível de importância de classificar e reutilizar dimensões em BI a partir de conceitos e relações em AE?

Qual o nível de importância de classificar e reutilizar métricas em BI a partir de KPI e objetivos em AE?

	Não é relevante	0	0%
Não é possível pois em AE não se consegue ter todas as formulas de cálculo como regras de negócio		1	10%
	Otimizaria o esforço de relacionar KPI em AE com métricas em BI	8	80%
	Other	2	20%

Figura 5.21: Qual o nível de importância de classificar e reutilizar métricas em BI a partir de KPI e objetivos em AE?

Na Figura 5.22 é apresentado o resultado sobre a utilização de linguagem natural para facilitar a criação de uma linguagem organizacional suportada em metadados como ontologia. Neste caso, 80% dos inquiridos considera que otimizaria a forma de pesquisa para utilizadores menos “tecnológicos”.

Qual o nível de importância de Linguagem Natural como interface com utilizador em AE e BI?

	Não é relevante	2	20%
Otimiza a forma de pesquisa para utilizadores menos “tecnológicos”		8	80%
	Other	0	0%

Figura 5.22: Qual o nível de dificuldade na gestão de *business intelligence*?

6 Arquitetura proposta

Neste capítulo apresenta-se a arquitetura de solução proposta considerando os requisitos derivados do problema (sub-capítulo 6.2), os cenários alternativos de solução (sub-capítulo 6.3) e a arquitetura de solução (sub-capítulo 6.4).

6.1 Introdução

Tendo por base o problema de interoperabilidade entre arquitetura empresarial e *business intelligence*, foram considerados dois cenários de solução como análise preliminar para se poder focar a investigação:

- **Cenário 1:** modelo comum representado num repositório partilhado entre os dois sistemas em estruturas de metadados (definição de conceitos e relações);
- **Cenário 2:** modelo de metadados criado como uma ontologia, por integração de vários conceitos entre arquitetura empresarial, bases de dados e glossários, permitindo a sua utilização para descoberta de conhecimento em metadados, mas sem integração em cada sistema.

O cenário 1 tem no entanto os seguintes constrangimentos:

- Existe uma norma ao nível de linguagem de notação para arquitetura empresarial (ArchiMate) mas nem todas as ferramentas de AE a suportam pois algumas utilizam modelos proprietários;
- Não existe uma norma de *business intelligence* com estruturas de dados e *dashboard*, nem como metamodelo nem como detalhe por sector de atividade;
- Existe uma norma para metadados transversal (IMM), mas com pouca adesão atual de ferramentas ao nível de aplicações de suporte ao negócio, ferramentas de AE e ferramentas de BI;
- As tecnologias atuais, de arquitetura empresarial ou de *business intelligence*, utilizam modelos proprietários. Como tal, a solução implicava passagens de definições entre os sistemas, com impacto na transformação e atualização de metadados. Este tipo de solução resultaria em implementação pouco realista visto que implicaria que cada sistema tivesse que implementar este tipo de interface, que é o que até hoje se tem tentado com metamodelos comuns, mas com pouca adesão;

Como tal, para que este cenário 1 fosse uma solução adequada, teria que se utilizar um sistema que permitisse modelar a arquitetura empresarial e os metadados para o sistema de *business intelligence* reutilizando os mesmos metadados ou permitindo a conversão entre os metadados de cada sistema, o que implicava serem modelos com mecanismos expansíveis para importação e exportação de metadados. De qualquer forma, para se avaliar este tipo de abordagem ter-se-ia que avaliar a adequação do conceito de alinhamento e respetivo impacto. Por essa razão, focou-se a investigação na avaliação do problema crítico que corresponde ao cenário 2, alinhada com a questão central da investigação, isto é, será adequado utilizar metadados como ontologias para criar interoperabilidade entre sistemas de arquitetura empresarial e sistemas de *business intelligence*?

6.2 Requisitos derivados do problema

Face ao problema identificado no capítulo 4, foram detalhados os seguintes requisitos para se encontrar uma solução:

- Definir na arquitetura empresarial as “áreas temáticas de informação”³ (e.g. orgânica, negócio, custos, proveitos) e para cada área os principais conceitos e suas relações;
- Definir em *business intelligence* o modelo de dados, modelo de integração de dados e modelo de visualização de dados, que permitam um alinhamento com as áreas temáticas de informação definidas na arquitetura empresarial, detalhando os conceitos relacionados com artefactos de tabelas, campos, dimensões e métricas em *business intelligence*;
- Integrar em ontologia com especificação OWL os modelos de conceitos da arquitetura empresarial, glossário, tabelas e campos de bases de dados, para criar uma ontologia para uma entidade (ontologia de domínio);
- Permitir analisar expressões de informação de gestão com uma estrutura padrão de “ANALISAR <métricas> DE <domínios de informação> POR <dimensões> PARA <restrições>”, garantindo a decomposição da frase através de linguagem natural para identificar conceitos (palavras) relacionadas com artefactos de metadados integrados na ontologia definida.

Entendendo o enquadramento do problema, foram considerados os seguintes cenários alternativos de solução para resposta aos requisitos, tal como indicado na Tabela 6.1:

Cenário	Descrição	Vantagem	Desvantagem
1.AE e BI com mesmos metadados.	Juntar os dois sistemas num único, com metadados centrais, ou criar interfaces entre os dois repositórios de metadados.	Permite reutilizar os mesmos conceitos reduzindo o impacto de alterações pois são definidos uma única vez.	Teria que ser o mesmo sistema (que não existe) com um único repositório de metadados. Senão obrigaria a criar interfaces complexas de conversão entre metamodelos que são normalmente fechados.
2.AE como base de ontologia sem integração direta em BI.	Utilizar AE para definir a ontologia inicial, enriquecida com base de dados e glossários.	Permite separar o desenho do modelo da organização como uma ontologia da sua instância com elementos detalhados, criando-se uma solução de metadados corporativos como ontologia. Evita desalinhamento pois a ontologia funciona como modelo intermédio de definições relacionadas.	Não permite a reutilização direta por cada ferramenta.

Tabela 6.1: Cenários alternativos de hipótese de solução

³ Tradução do termo em inglês “Subject Area” no domínio de *business intelligence*

Face à análise de vantagens e desvantagens, optou-se pelo cenário 2. A opção é reforçada pelo facto de ser uma investigação em novos domínios, sem referências equivalentes e estruturas que já existam para se montar soluções adicionais. Apesar de existirem modelos como o IMM (2008), apresentado no sub-capítulo 3.3.3.2, com uma centralidade face a outros modelos, na realidade não existem ferramentas que suportem este metamodelo que permita testar a interoperabilidade. Considerando os requisitos derivados do problema, podemos considerar o seguinte para efeito de implementação de um protótipo para experimentação:

- **Requisito:** definir na arquitetura empresarial as áreas temáticas de informação (e.g. orgânica, negócio, custos, proveitos) e para cada área os principais conceitos e relações:
 - Criar os objetos de metamodelos ao nível de *business layer* para metadados de negócio, que são relevantes para detalhar a arquitetura, com visualizações para “orgânica/custos” e “negócio/proveitos”, criando relação entre artefactos;
 - Implementar os objetos de metamodelo ao nível de *application layer*, para metadados técnicos, que são relevantes para criar a ontologia ao nível de relação entre tabelas, campos, glossário e conceitos de arquitetura empresarial.
- **Requisito:** definir em *business intelligence* o modelo de dados, modelo de integração de dados e modelo de visualização de dados, que permitam um alinhamento com as áreas temáticas de informação definidas na arquitetura empresarial:
 - Criar em ETL as áreas de integração de dados de acordo com as áreas temáticas de arquitetura empresarial;
 - Desenhar tabelas e campos tendo por base as definições de conceitos utilizados na arquitetura empresarial, com uma indicação de dimensões e métricas por interpretação da ontologia;
 - Implementar *dashboard* criando perspetivas de análise por áreas temáticas definidas na arquitetura empresarial e considerando as dimensões e métricas identificadas como requisitos através de expressões de interrogação de negócio.
- **Requisito:** integrar em ontologia OWL os modelos de conceitos de arquitetura empresarial, glossário, tabelas e campos de bases de dados de *business intelligence*:
 - Permitir configurar contexto de ontologia com *namespace*, ficheiros de configuração e tipologia de ontologia;
 - Permitir guardar uma ontologia em base de dados para posterior utilização;
 - Carregar a ontologia inicial a partir de definição de conceitos de arquitetura empresarial criada em em ArchiMate e gerada para OWL;
 - Complementar a ontologia, com as definições de base de dados a partir de DDL gerado de uma base de dados de informação de gestão;
 - Complementar a ontologia previamente carregada, com as definições de glossários, a partir de um ficheiro Microsoft Excel onde para cada campo se pode associar conceitos associados;

- **Requisito:** permitir analisar expressões de informação de gestão com uma estrutura padrão de “analisar <métricas> de <domínios de informação> por <dimensões> para <restrições>, garantindo a decomposição da frase em linguagem natural para artefactos de metadados.
 - Permitir gravar expressões já analisadas e recupera-las para posterior utilização;
 - Analisar a expressão, criando as interpretações entre termos utilizados na expressão em linguagem natural e os respetivos termos técnicos nos metadados, tendo por base a ontologia criada.

6.3 Arquitetura de solução

6.3.1 Arquitetura de referência

Considerando os requisitos definidos, a arquitetura de solução considera o seguinte:

- Criação de *templates* para definição de uma arquitetura empresarial e sistema *business intelligence* para serem reutilizados como parte de uma abordagem metodológica de construção de sistemas de *business intelligence* a partir de requisitos definidos como parte do modelo da organização representado na arquitetura empresarial;
- Implementação de um protótipo de metadados suportado em ontologia OWL, que permite classificar as métricas e dimensões, além de permitir analisar de forma holística os conceitos chave da organização. Este mesmo modelo de ontologia suporta uma interrogação por linguagem natural, para o idioma Português, delimitado para uma estrutura de frase utilizada de forma típica em informação de gestão.

No caso do protótipo de metadados suportado em ontologia OWL, a arquitetura de referência para a solução apresentada na Figura 6.1, considera o seguinte:

- O modelo de arquitetura empresarial enquanto base para a ontologia, foi criado com o metamodelo e linguagem ArchiMate na ferramenta *freeware* Archi versão 2.7.1. Esta solução é descrita no sub-capítulo 6.3.2;
- O sistema de *business intelligence* foi criado com a tecnologia Microsoft BI (SQLServer, Integration Services e PowerBI Desktop), em regime *freeware*, o que apresenta limitações no funcionamento dos produtos, mas sem constrangimentos para o que se pretende investigar. Esta solução é descrita no sub-capítulo 6.3.3;
- Tendo por base o modelo de arquitetura empresarial criado em Archi e gerado para ontologia em OWL, é necessário consolidar com bases de dados e glossários. Para tal, utiliza-se ArchiOWL (transforma modelos ArchiMate em OWL), ApacheJENA (OWL, inferência e SPARQL) e OpenNLP (processamento de língua natural). Estas várias componentes foram integradas numa solução desenvolvida em Java com interface Swing. Esta solução é descrita no sub-capítulo 6.3.4.

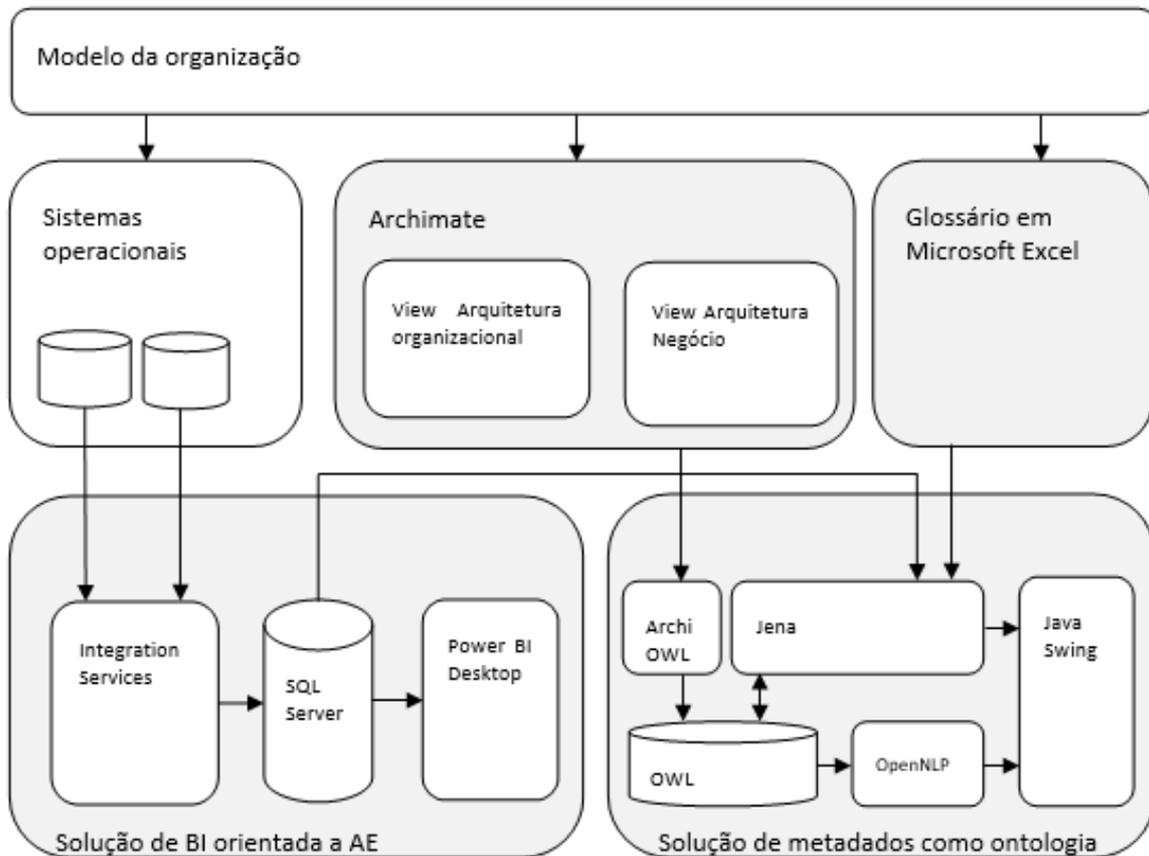


Figura 6.1: Arquitetura de referência de hipótese de solução

6.3.2 Metadados como ontologia

Os metadados como ontologia foram conceptualizados considerando a necessidade de integrar os modelos de *business intelligence* (bases de dados, ETL e *dashboard*) e arquitetura empresarial (conceitos e glossários), em torno de metadados técnicos e metadados de negócio, interrelacionados, tal como apresentados na Figura 6.2, destacando-se o seguinte:

- Cada tipo de componente tem uma hierarquia de objetos. Tal é o caso de ETL (*package, data, transformation*), base de dados (base dados, tabela, campo, formato) e relatórios (relatório, *query*);
- Os campos das bases de dados são o objeto central, pois permitem a relação com os relatórios, ETL e com a arquitetura empresarial. Podem ser classificados como métricas e dimensões a partir da sua nomenclatura (e.g. Cod, Desc), formato (e.g. *decimal, string*) ou herdarem a classificação de conceitos associados à arquitetura empresarial (pelo tipo de elemento) ou ao nível de glossário (morfologia de palavras de significado);
- A relação entre metadados técnicos e metadados de negócio surge através de campos e tabelas relacionados com a implementação de conceitos da arquitetura empresarial que são por sua vez vistos segundo significados normalmente definidos em glossários.

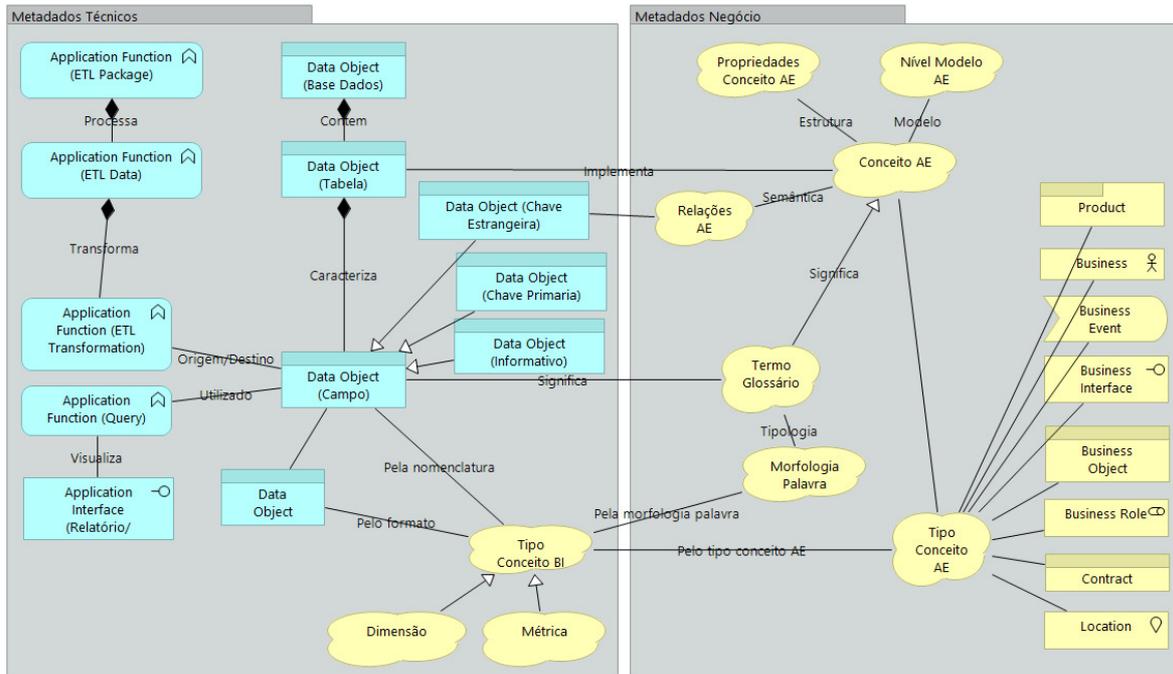


Figura 6.2: Modelo hipótese de metadados como ontologia

Para a sua implementação, considerou-se o seguinte:

- O modelo de arquitetura empresarial em ArchiMate, utilizando a ferramenta Archi, tem a possibilidade de exportar para OWL através de um *plug-in* desenvolvido por Szwed (2015), sendo que considera como classes os níveis (e.g. *business layer*, *application layer*) e os elementos em cada nível (e.g. *actor*, *business object*). Por outro lado, considera como relações não só as hierarquias de classes entre níveis e elementos em cada nível, mas também as *structural relationship* entre os elementos (e.g. *association*, *access*, *used by*). Os indivíduos correspondem aos elementos e suas relações;
- O glossário permite criar relação entre cada elemento de arquitetura empresarial e termos equivalentes (e.g. oferta é sinónimo de produtos), além de possibilitar caracterizar os campos em termos de descrição (e.g. codCliente é sinónimo de cliente), o que permite dotar a ontologia de significado para cada conceito;
- O modelo de bases de dados é exportável para DDL, com a definição de cada tabela e seus campos dentro de uma base de dados e respetivos formatos. Neste caso, consideramos que cada tabela e campo correspondem a classes. As relações entre campos e tabelas utilizam a designação “caracterizado por”. A relação entre tabelas utiliza a designação equivalente ao nome da chave estrangeira;
- A partir do formato dos campos e sua nomenclatura, classifica-se cada campo como métrica e dimensão pelo nome dos campos face a um padrão (e.g. Cod, Desc) e formato (e.g. *numeric*, *string*). Por outro lado, cada conceito, de acordo com a sua análise

gramatical, foi classificado por NLP como métrica ou dimensão, permitindo por inferência que os campos e tabelas associados fiquem com a mesma definição.

6.3.2.1 Metadados técnicos

Para criar um modelo de metadados técnicos que considere de forma integrada as bases de dados, glossários e conceitos de arquitetura empresarial, estendeu-se o modelo definido por Astrova (2009), tal como apresentado na Figura 6.3, ajustando-se os nomes e criando novos elementos e utilizando-se as designações definidas em ArchiMate [e.g. *data object* (Base Dados)].

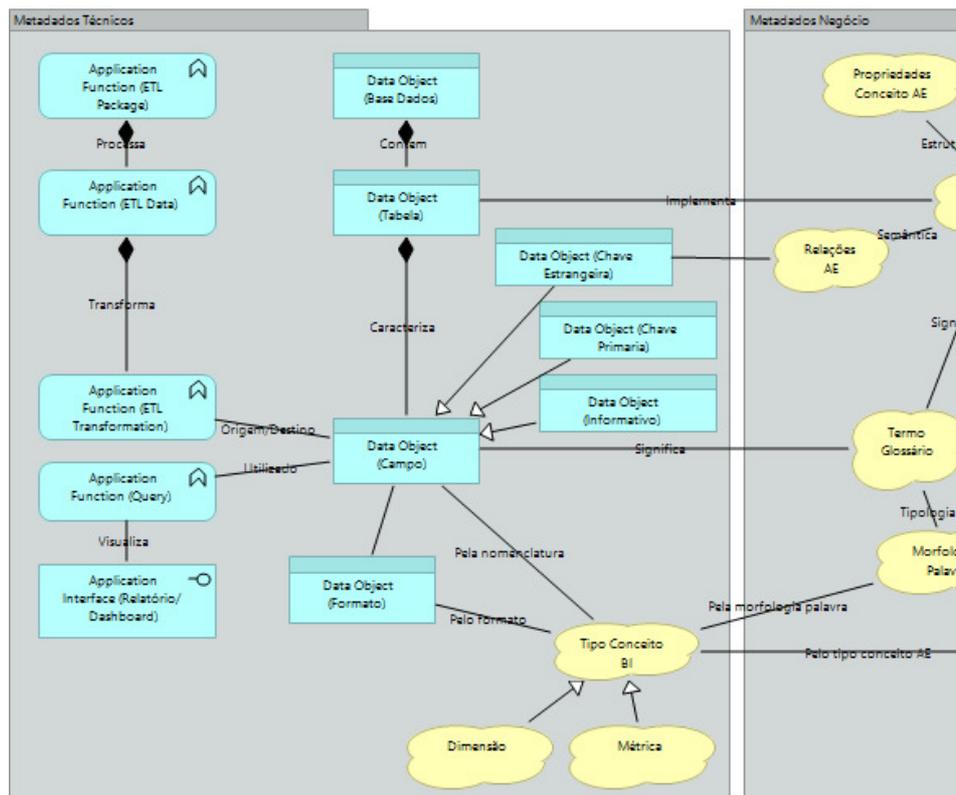


Figura 6.3: Metadados técnicos

Este modelo é caracterizado pelo seguinte:

- As tabelas de bases de dados têm uma classificação enquanto tabelas de factos ou dimensões a partir da sua extensão de definição enquanto conceito de arquitetura empresarial e conceitos equivalentes definidos num glossário;
- Os campos das tabelas, têm classificações enquanto métricas e dimensões, além de definições críticas para criação de ontologias ao nível de tipo de campo (chave primária, chave estrangeira, campo informativo). Esta classificação depende da extensão de definição enquanto conceito de arquitetura empresarial e conceitos equivalentes definidos num glossário, além de que ao nível do “Tipo Conceito BI”, os campos são igualmente classificados a partir da sua nomenclatura (e.g. Cod<nome>, Desc<nome>) e formato (e.g. *decimal*, *string*, *numeric*, *date*);

- Os campos entram em *queries*, que por sua vez são utilizadas em relatórios/dashboard para visualização da informação. Os campos entram igualmente em transformações ao nível do ETL;
- O ETL é composto por packages que permitem transformar tabelas a partir de regras de negócio de movimentação de dados em campos entre origem e destino;
- Os relatórios/*dashboard* são objetos de visualização que utilizam *queries* que por sua vez utilizam campos. No caso do BI, esses campos são normalmente vistos como dimensões ou métricas.

Ao nível de AE estes conceitos podem ser implementados de acordo com o descrito na Tabela 6.2, com conceitos ArchiMate do nível de arquitetura *application*, onde se pode notar que existem tipos de objetos que são reutilizados por falta de objectos específicos.

Conceito ArchiMate	Descrição	Classificação BI
<i>Application Interface.</i>	Componente de visualização de dados.	Dashboard/Relatório.
<i>Application Function.</i>	Componente DML para acesso a campos de várias tabelas.	<i>Query</i> de exploração de dados.
<i>Application Function.</i>	Componente agregados de transformações, designado por <i>package</i> em Microsoft <i>Integration Service</i> , ou pode ser visto igualmente como cadeia Batch.	<i>ETL Package.</i>
<i>Application Function.</i>	Componente para transformação de uma tabela específica enquanto destino dos dados, designado por <i>data flow task</i> em Microsoft <i>Integration</i> ou pode ser visto genericamente como Job.	<i>ETL Data.</i>
<i>Application Function.</i>	Componente de transformação entre campos e tabelas origem e destino. Designado por <i>task</i> em <i>Microsoft Integration Services</i> (<i>source</i> , <i>destination</i> ou <i>transformation task</i>) mas pode ser visto genericamente como Step. Pode-se compor sequências de <i>tasks</i> para criar um fluxo de processamento entre a leitura de dados, transformação e carregamento. Em cada <i>task</i> pode-se utilizar regras de transformação (e.g. Saldo Crédito = Capital + Juros + Impostos + Despesas).	<i>ETL Transformation.</i>
<i>Data Object.</i>	Conjunto de tabelas de um modelo de dados.	Base de dados.
<i>Data Object.</i>	Tabela com vários campos.	Tabela.
<i>Data Object.</i>	Campo de dados que pode ser uma chave primária, chave estrangeira ou campo informativo.	Campo.
<i>Data Object.</i>	Formato genérico de um campo (e.g. <i>string</i>), que pode ser ajustado por cada campo (e.g. <i>string</i> [20]).	Formato campo.

Tabela 6.2: Conceitos de metadados técnicos

6.3.2.2 Metadados de negócio

Para criar um modelo de metadados de negócio, partiu-se de uma generalização de conceitos da arquitetura empresarial que têm propriedades, relações e que existem no modelo *business layer* do modelo ArchiMate. No entanto, utiliza-se a especialização para detalhar os elementos ArchiMate utilizados (e.g. *actor*, *business object*), que permitem implementar modelos concretos com conceitos de negócio (e.g. cliente, canal, produtos). Tendo por base a modelação de arquitetura empresarial, estendeu-se a mesma com o glossário de termos e conceitos da organização. Por esta razão se centra a modelação da arquitetura de negócio no conceito “Termo Glossário” e “Conceito AE”, apresentado na Figura 6.4.

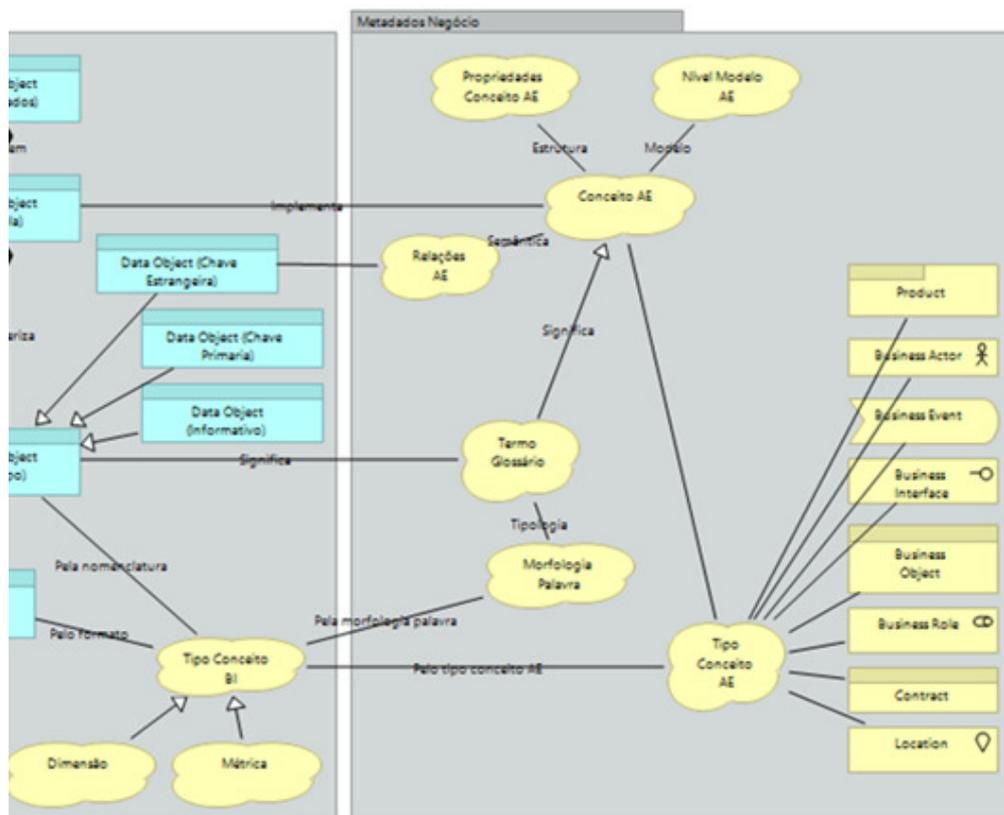


Figura 6.4: Metadados de negócio

Como tal, considerou-se a sistematização do problema no sub-capítulo 4.2, onde se enquadra as áreas temáticas de análise “orgânica” e de “negócio”, que estruturam as dimensões e métricas base, reutilizadas depois numa análise específica de custos, proveitos e rentabilidade. Neste sentido, tendo por base o modelo ArchiMate, considerou-se as seguintes definições base de acordo com GROUP (2012, 2016) para os seguintes tipos de conceito:

- *Active structure* são entidades capazes de ter determinados comportamentos (isto é, *behavioral*). Correspondem ao sujeito quando se analisa uma frase;
- *Passive structure* são objetos sobre os quais os comportamentos atuam. Correspondem ao objeto quando se analisa uma frase;
- *Behavioral* são as atividades executadas pelos elementos *Active Structure* sobre os elementos *Passive Structure*. Correspondem ao verbo quando se analisa uma frase.

Tendo por base estas definições, foram analisados os vários conceitos de acordo com o metamodelo ArchiMate, para se definir os objetos adequados a utilizar na modelação da arquitetura orgânica/custos e arquitetura negócio/proveitos, bem como a sua relação com BI, tal como apresentado na Tabela 6.3.

Conceito ArchiMate	Tipo conceito ArchiMate	Tipo arquitetura		Descrição	Classificação BI
		Orgânica	Negócio		
<i>Actor.</i>	<i>Active Structure.</i>	X	X	Representa uma entidade. Pode ser uma unidade orgânica, fornecedor, colaborador ou cliente.	Dimensão.
<i>Role.</i>	<i>Active Structure.</i>	X	X	Representa uma responsabilidade. Pode ser uma função de um colaborador ou um segmento de cliente.	Dimensão.
<i>Location.</i>	<i>Active Structure.</i>	X		Representa um local no espaço geográfico. Corresponde a locais onde a organização tem atividade.	Dimensão.
<i>Business Interface.</i>	<i>Active Structure.</i>		X	Representa um ponto de acesso onde os serviços são disponibilizados. Corresponde a canais comerciais.	Dimensão.
<i>Contract.</i>	<i>Passive Structure.</i>	X	X	Representa uma forma de acordo, com direitos e obrigações relacionado com produtos e serviços. Pode ser um contrato de custos ou de proveitos.	Fonte de métricas.
<i>Business Object.</i>	<i>Passive Structure.</i>		X	Representa algo que tenha significado para o negócio.	Dimensão.
<i>Business Event.</i>	<i>Behavioral.</i>	X	X	Representa algo que ocorre e que influencia ou resulta de comportamentos da organização. Corresponde a transações.	Fonte de métricas.
<i>Business Service.</i>	<i>Behavioral.</i>		X	Representa um serviço disponibilizado aos clientes. Corresponde à oferta comercial.	Dimensão.
<i>Business Process.</i>	<i>Behavioral.</i>	X	X	Representa um grupo de atividades com entradas, processamento e saída para produzir produtos ou serviços. Corresponde a formas de tratamento de transações para a sua realização.	Dimensão.

Tabela 6.3: Conceitos de metadados de negócio

Com base nesta definição, foram definidos metamodelos para a arquitetura organizacional e arquitetura de negócio, tal como apresentado na Figuras 6.5 e 6.6.

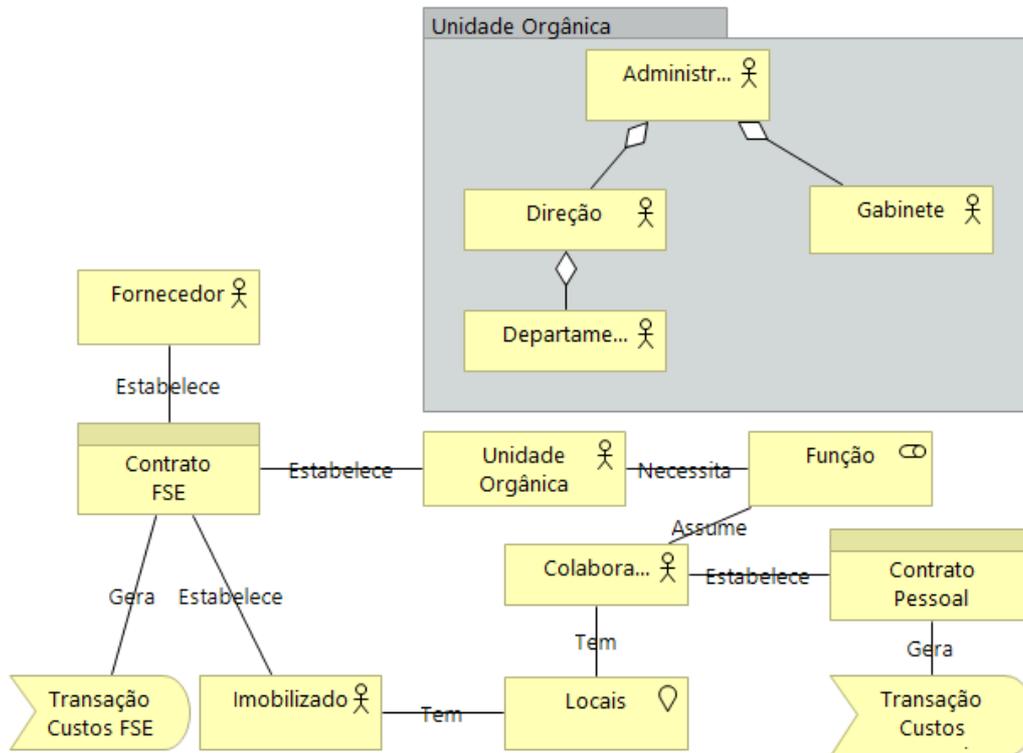


Figura 6.5: Ontologia para arquitetura organizacional

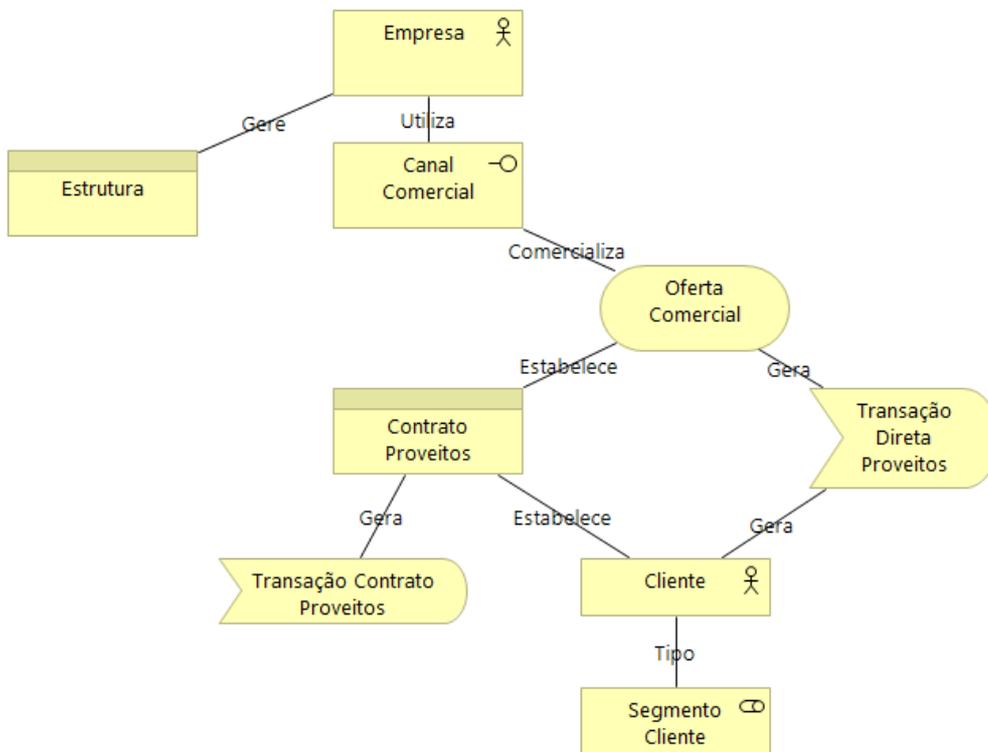


Figura 6.6: Ontologia para arquitetura de negócio

Para cada modelo, é criado um modelo de relação com a arquitetura de informação, tal como apresentado na Figura 6.7, onde se relaciona cada conceito com tabelas críticas. De notar que são consideradas somente as tabelas, pois os campos são relacionados com as tabelas a partir do DDL e por sua vez o significado é obtido via glossário para cada termo (conceito AE ou campo) e com vários significados.

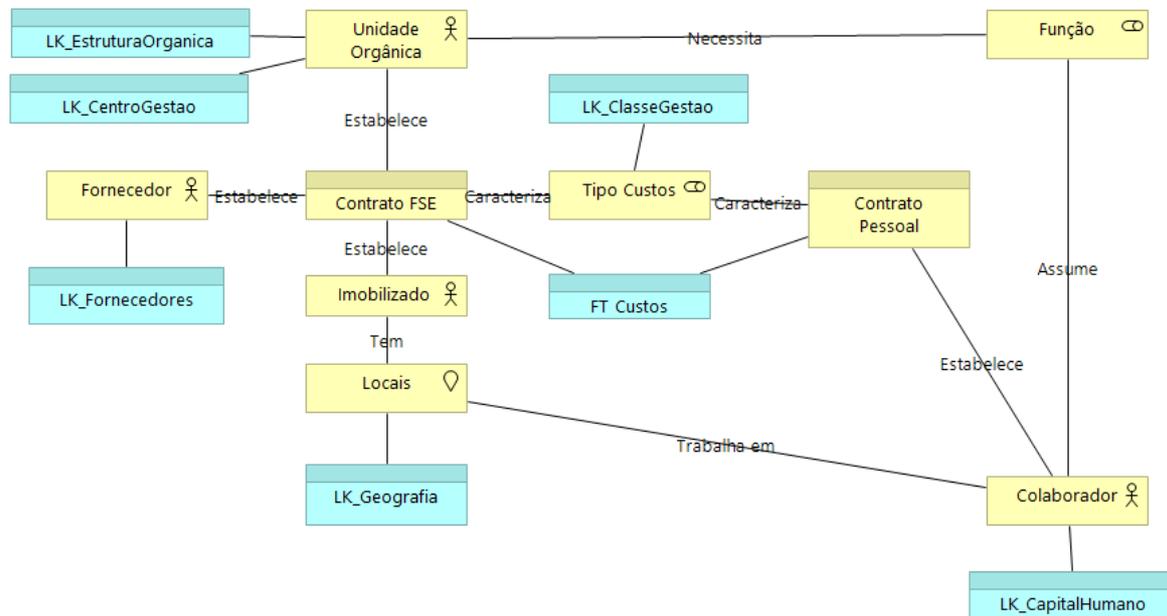


Figura 6.7: Modelo de relação entre arquitetura negócio/organizacional e arquitetura de informação

Estes modelos, vistos com ontologia, permitiram identificar os conceitos chave para cada área de informação, orgânica/custos e negócio/proveitos, para que ao nível de *business intelligence* se possa definir os modelos de dados, processos de integração e formas de exploração da informação da organização.

6.3.3 Solução de *business intelligence* orientada a arquitetura empresarial

Para se implementar uma solução de *business intelligence*, definiu-se uma abordagem, apresentada na Figura 6.8 que considere a arquitetura empresarial como fonte inicial de definição de requisitos enquanto dimensões e métricas, mas igualmente ao nível de áreas de informação que determinam como são organizados os artefactos de *business intelligence*.

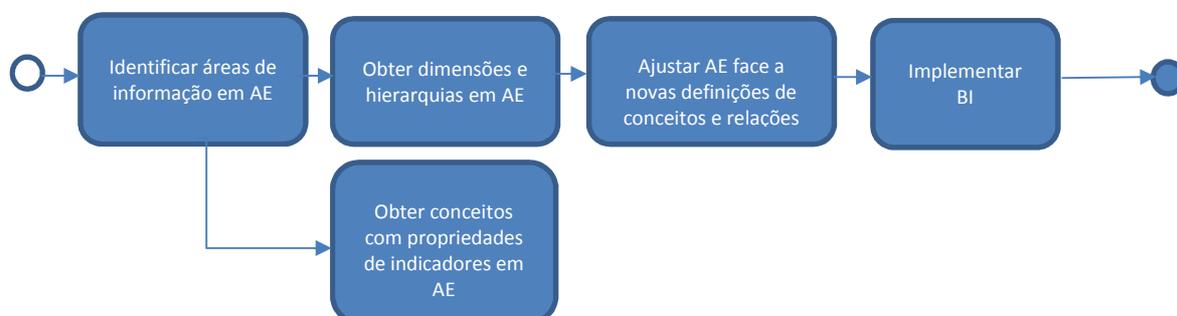


Figura 6.8: Abordagem de implementação de *business intelligence*

Ao nível de processos de integração em *business intelligence*, foram considerados somente princípios de estruturação de processos considerando que as ferramentas de ETL têm por base os seguintes componentes:

- Forma de agregar *Jobs* por áreas, que neste caso correspondem aos níveis de arquitetura, exceto casos transversais entre vários níveis, como é o caso de tabelas de tempo ou geografia. Na Figura 6.9 é apresentado um caso em Microsoft *Integration Services* onde se estruturam 2 Packages (Contexto e Operacional), sendo que no caso do *package* “Operacional”, se coloca agregação de *jobs* por áreas através de *sequence container* (e.g. Orgânica), em que cada um se coloca os Jobs para carregamento de tabelas específicas (e.g. LK_EstruturaOrganica);
- *Job* que agrega um conjunto de tarefas, sendo que cada tarefa tem normalmente um objeto que representa a origem da informação, um objeto que representa a transformação e um objeto que representa o carregamento no destino da informação. Neste caso, cada *Job* corresponde a uma tabela, isto é, conceito do modelo de ontologia a tratar. Na Figura 6.10 é apresentado um caso em Microsoft *Integration Services* em que um *job* visto como *control flow*, tem um *data flow* (LK_EstruturaOrganica), com um componente que indica que a origem de dados está em Microsoft Excel, um processo de transformação (*script component*) e um objeto que representa a tabela destino (*ADO Net Destination*).

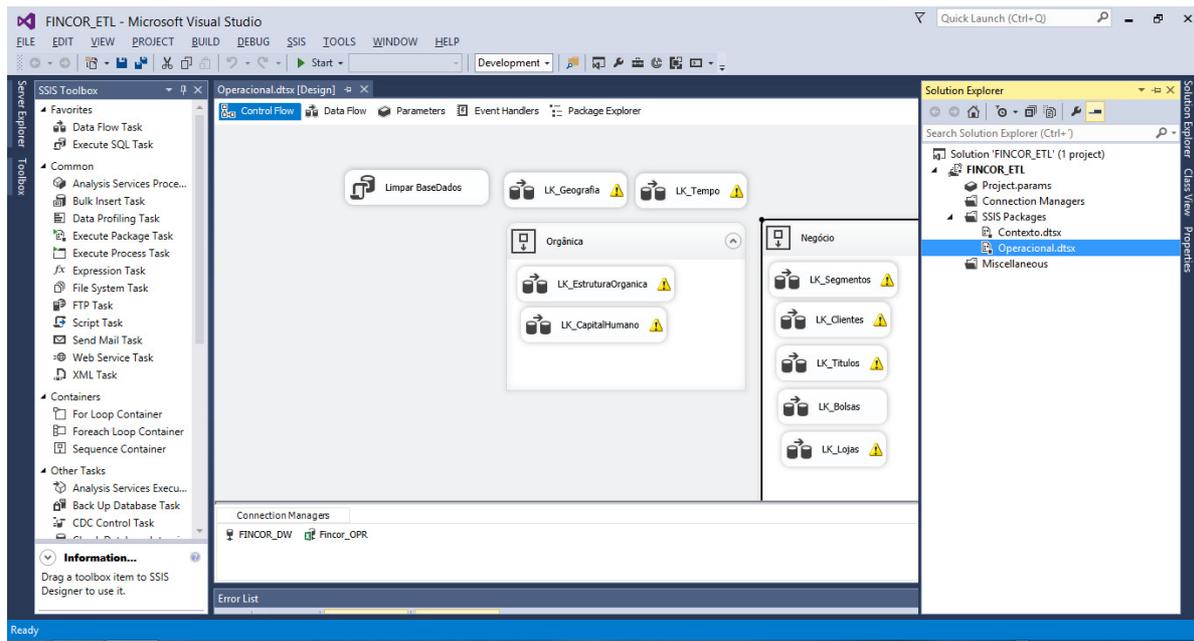


Figura 6.9: Agregação de Jobs para integração de dados em Microsoft *Integration Services*

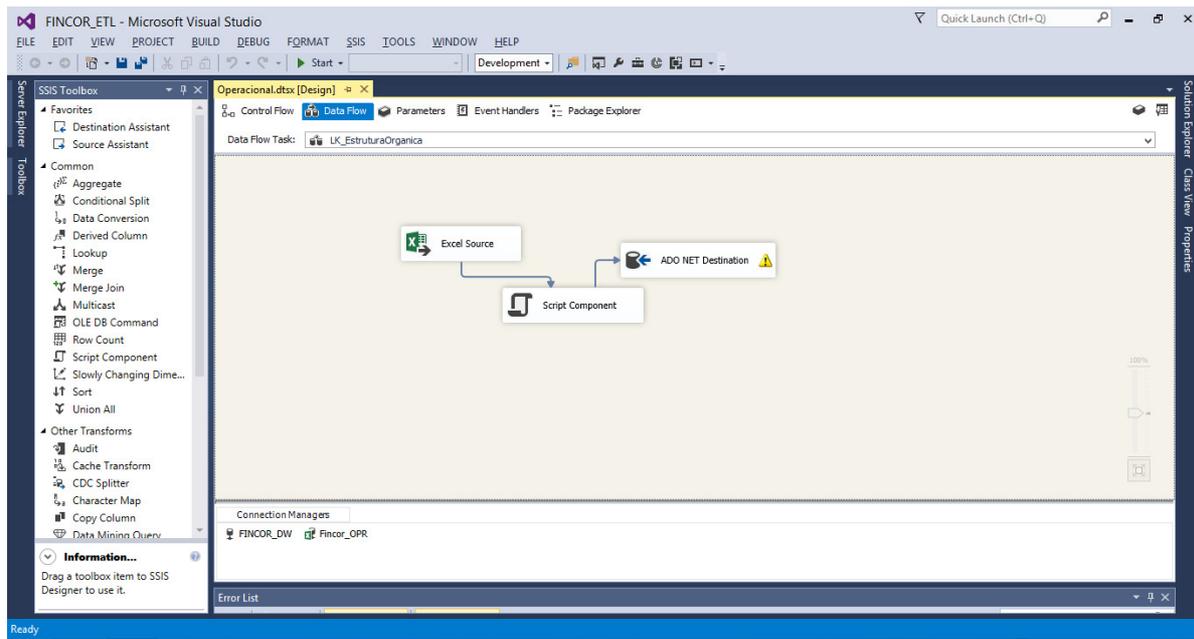


Figura 6.10: Processo de integração de dados de uma tabela em Microsoft *Integration Services*

Ao nível de dados, foram considerados modelos implementados em bases de dados com as seguintes características e representado na Figura 6.11 no caso de uma base de dados em Microsoft SQL Server, na Figura 6.12 no caso do modelo de dados para visualização e na Figura 6.13 com uma estrutura de relação entre tabelas em Microsoft Power BI:

- Tipo de modelação em *snow flake* tendo por base as hierarquias e sua relação com a estrutura de conceitos e respetivas relações com origem na ontologia definida;
- As tabelas com prefixo “FT” (Fact Table) correspondem a tabelas de factos onde as principais métricas e códigos de dimensões estão definidos, sendo que as dimensões são decodificadas em tabelas com prefixo “LK” (Lookup Table). As tabelas LK podem estar relacionadas entre si de acordo com hierarquias de negócio;
- As dimensões correspondem aos conceitos identificados na ontologia, respeitando as nomenclaturas e hierarquias definidas. Numa arquitetura empresarial corresponde a relações entre conceitos;
- As métricas são representadas como tabelas de factos de “Proveitos” e “Custos” que agregam todas as métricas agrupadas pelas dimensões definidas. As métricas têm origem em conceitos de eventos, que agregam várias métricas. Por exemplo, uma transação de proveitos pode indicar métricas de valor da fatura, valor cobrado, valor de impostos, entre outras rúbricas financeiras, além da contagem de faturas;
- Podem existir tabelas de factos complementares, desde que existam sempre os “Proveitos” e “Custos”. Tal é o caso de detalhe de tabelas de faturação (recebidas ou emitida), carteira de clientes e orçamentos, entre outras. No entanto, para focarmos a nossa investigação, consideramos somente os "Custos" e “Proveitos”;
- A relação entre tabelas é mantida, como um modelo semântico, nas ferramentas de *business intelligence* e não no sistema de gestão de base de dados.

No caso da Figura 6.11 apresenta-se a base de dados em Microsoft SQLServer com as respetivas tabelas sendo que as tabelas foram criadas com script DDL gerado a partir de Microsoft Integration Server por ligação direta às fontes de dados mas ajustando as nomenclaturas antes da criação das tabelas.

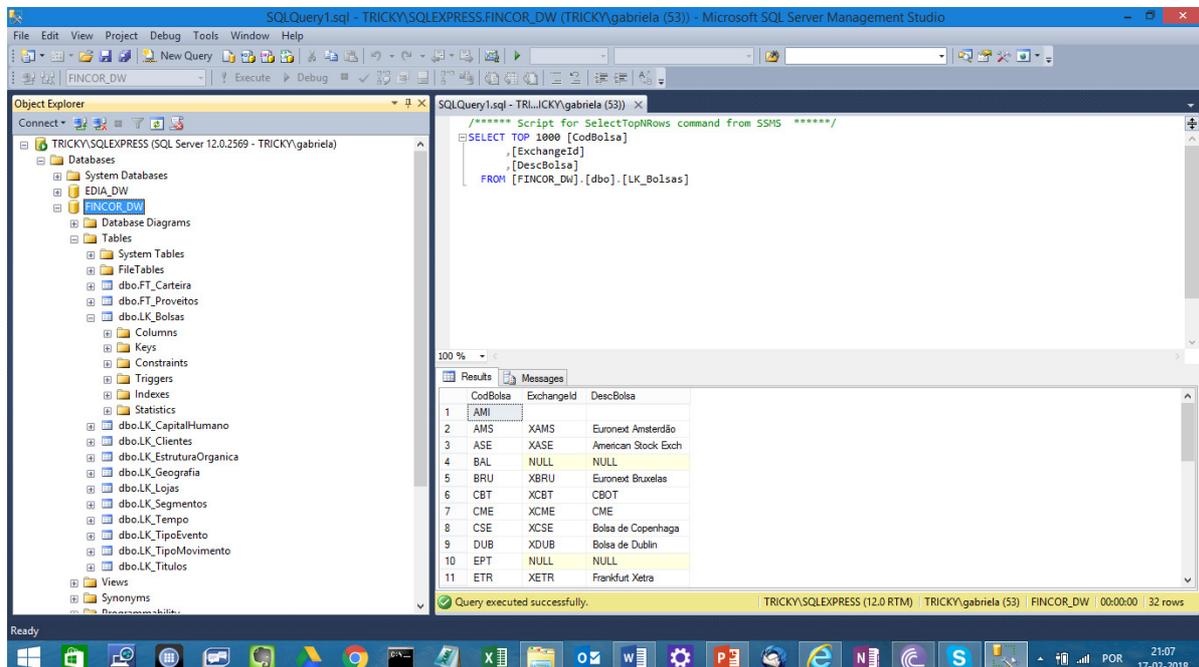


Figura 6.11: Base de dados em Microsoft SQLServer

No caso da Figura 6.12 apresenta-se o modelo de dados criado em Microsoft PowerBI diretamente com as respetivas relações entre as chaves considerando as chaves estrangeiras, sendo que a ferramenta deteta as relações de acordo com a nomenclatura dos campos mas permitindo acertar na cardinalidade ou na criação de relações específicas que seja necessário ajustar, como é mostrado na Figura 6.13.

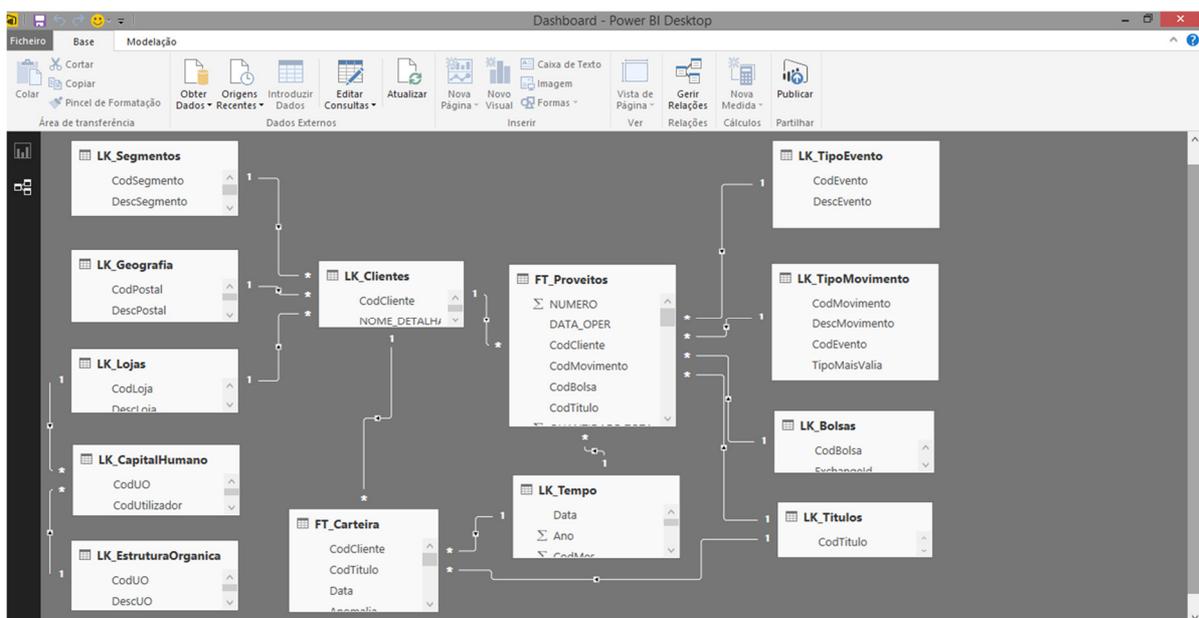


Figura 6.12: Modelação de dados em Microsoft PowerBI (Tabelas)

Editar Relação

Selecione as tabelas e colunas que se relacionam entre si.

FT_Carteira

CodCliente	CodTitulo	Data	Anomalia	Dívida	Cambio	ValorUnitario	QtdSaldo	Val
100002	IVV US	31/12/2014 00:00:00	null	null	0	0	9040	
100002	LYXIB SM	31/12/2015 00:00:00	null	EUR	1	94,12	-3092	
100053	AEP US	31/12/2014 00:00:00	null	null	0	0	-46	

LK_Clientes

CodCliente	NOME_DETALHADO	EMAIL	Morada	MORADA_1	MORADA_2	MORADA_3	C
100114	BARCLAYS BANK PLC - 041	null	null	null	null	null	00
300072	Ferox Capital LLP	null	null	66 St James's Street,	null	null	SV
100100	Ox Global Investments Ltd.	null	null	11 John Princes Street	null	null	W

Cardinalidade: Muitos para Um (*:1)

Direção de filtro cruzado: Único

Tornar esta relação ativa

Assumir Integridade Referencial

[Mais Informações](#)

OK Cancelar

Figura 6.13: Modelação de dados em Microsoft PowerBI (Relações entre tabelas)

Ao nível de *dashboard* em *business intelligence*, considerou-se a representação em quatro áreas temáticas, vistas como áreas de representação de conhecimento, relacionadas com “organização”, “negócio”, “custos” e “proveitos”, com as seguintes características, tal como apresentado na Figura 6.14:

- Cada área reutiliza os conceitos da arquitetura empresarial enquanto dimensões e métricas, com os mesmos nomes;
- A navegação por hierarquias e relações entre dimensões deve ser feita por conceitos de “*drill*”, permitindo não só mostrar a ontologia (e.g. segmentos tem clientes) mas também o detalhe de métricas instanciado a partir de quantidade de clientes por dimensão e a desagregação por hierarquia da dimensão. No caso de informação muito detalhada é uma vantagem face a visualizações na arquitetura empresarial. O “*drill*” permite uma navegação entre hierarquias (e.g. de Segmento de cliente para cliente), recalculando as métricas em conformidade e por decomposição de dimensões. Este modelo permite mostrar a relação entre conceitos, com a vantagem de apresentar as métricas em conformidade com o nível de detalhe. Como tal, é uma alternativa mais adequada para visualização da informação, face a modelos de orgânica e relações visualizados numa arquitetura empresarial.

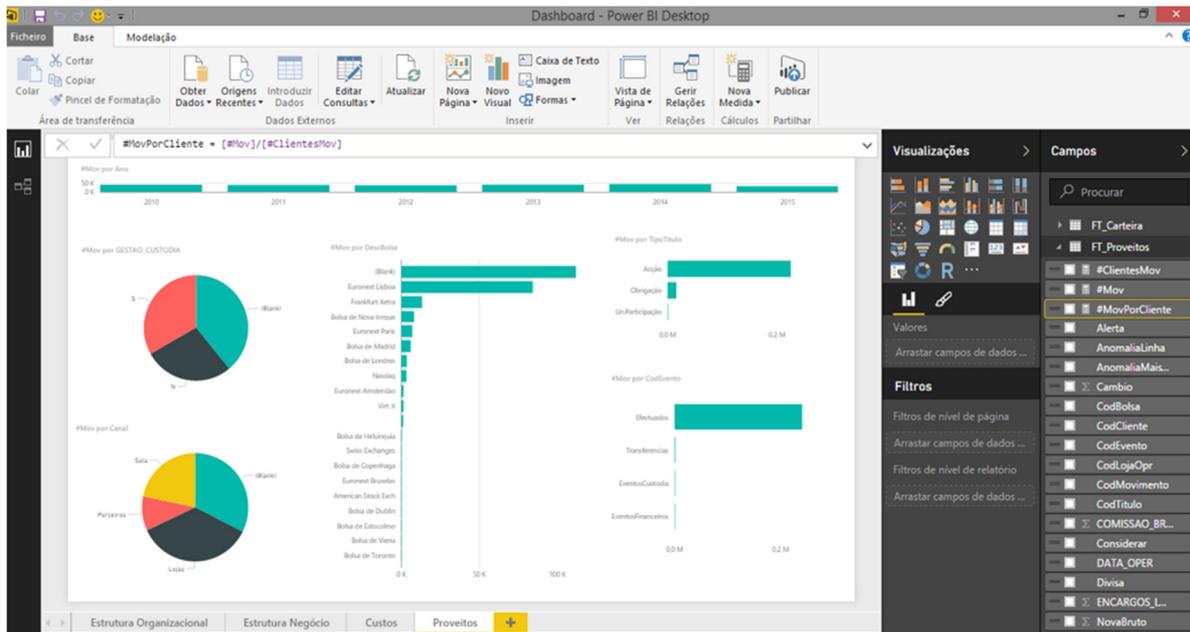


Figura 6.14: Estrutura de *dashboard business intelligence*

6.3.4 Solução de análise de metadados como ontologia

6.3.4.1 Arquitetura tecnológica

Para implementar o modelo de metadados como ontologia e analisador de expressões, foi montada uma arquitetura apresentada na Figura 6.15 e Tabela 6.4, em termos de componentes tecnológicos e respetivas interfaces que pode ser detalhada da seguinte forma:

- Os modelos de ontologia são capturados a partir de conversão de formatos ArchiMate (arquitetura empresarial) via Plugin ArchiOWL, glossários em Microsoft Excel/CSV e DDL de bases de dados obtidos via ligação JDBC a base de dados. Estes modelos são integrados via processamento de *match & merge* utilizando ApacheJENA via OWL e SPARQL, mas tendo como orientação a ontologia definida ao nível da arquitetura empresarial em ArchiMate. O modelo de ontologia inicial é o que for definido na arquitetura empresarial em Archi com modelo ArchiMate. Deve incluir não só os conceitos e relações chave (e.g. objetivos, produtos, processos, orgânica, pessoas) como também regras de relação específica ao nível de produtos (estrutura e equivalência de termos de acordo com o domínio de linguagem), canais (tipo e relação entre canais e com clientes) e clientes (segmentação de clientes);
- Primeiro integra-se o modelo gerado pela ferramenta Archi, para de seguida se integrar glossários para criar a relação “significa” com tabelas e campos, classificar cada termo pela sua morfologia e a sua classificação como métrica ou dimensão. Por fim, integra-se o DDL criando relação entre campos e tabelas, e para cada campo, classifica-se como métrica ou dimensões de acordo com o formato do campo e sua nomenclatura;
- Após a integração num modelo gera-se a inferência (*reasoning*) que por transitividade, permite criar indivíduos nas classes métricas e dimensões;

- Por fim, o modelo resultante é explorado por interface em Java Swing com suporte de OpenNLP para interface com utilizador em processamento de língua natural. Com o modelo e classificação efetuada, pode-se explorar a ontologia com base em linguagem natural com um padrão de linguagem de “Analisar <métricas> DE <área funcional ou domínio> POR <dimensões> PARA <condições ou restrições>”.

No caso do modelo de arquitetura na Figura 6.15, destaca-se a utilização de ApacheJENA para tratar de forma programada as ontologias de acordo com os processos de *ontology learning*, integrando várias fontes como é o caso de glossários via dados em Microsoft Excel, acesso direto a base de dados para obter o DDL, e o modelo base de ontologia criado em AE e gerado para OWL antes da integração. De notar igualmente o papel auxiliar de processamento de língua natural via Apache NLP para tratamento de expressões de texto.

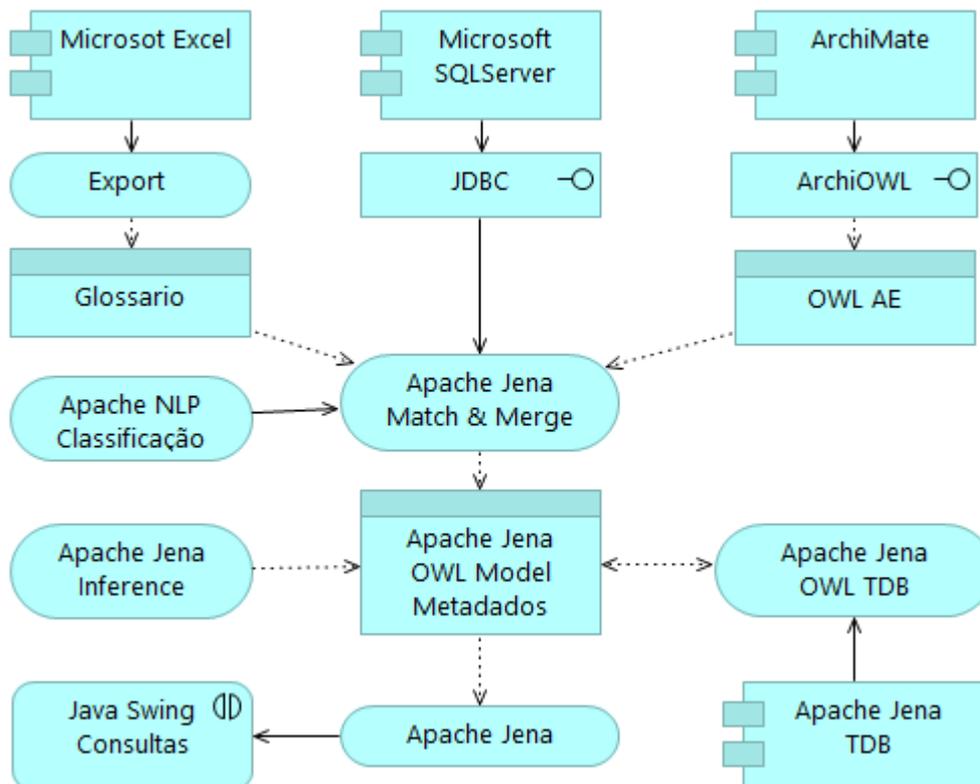


Figura 6.15: Arquitetura técnica de solução em hipótese

Os vários componentes da arquitetura são detalhados na Tabela 6.4 considerando o seguinte:

- Tecnologia utilizada, sendo que utilizamos tecnologia *open source* como é o caso de Apache que tem soluções para OWL e NLP;
- Forma de utilização, descrevendo-se a utilização feita de cada tecnologia;
- Motivo pelo qual se utiliza o componente.

Tecnologia	Descrição utilização	Motivação
Microsoft Excel.	Ferramenta utilizada para criar glossários ao permitir criar uma estrutura tabular de termo utilizado em AE e termos de glossário alternativos.	Poderia ser feito em ArchiMate, mas o Excel é normalmente utilizado nas empresas para documentar este tipo e informação, colocando-se alguma informação complementar sobre cada conceito.
Microsoft SQL Server.	Poderia ser qualquer SGBD, pois é necessário suporte para repositório de dados e <i>queries</i> DDL e DML.	
Archi.	Ferramenta de modelação de arquitetura empresarial com base no metamodelo ArchiMate e que permite instalar um Plug-In ArchiOWL para exportação do modelo ArchiMate para OWL.	Permitiu criar o modelo de ontologia base e os modelos de experimentação para cada caso de estudo para posterior conversão para OWL.
ArchiOWL.	Plug-in que é instalado em Archi para transformar um modelo em ArchiMate para OWL efetuando uma transformação de componentes e hierarquias para <i>class</i> OWL e gerando relações como <i>object properties</i> em OWL.	Permitiu reutilizar uma componente de conversão existente entre ArchiMate e OWL.
ApacheJENA.	<i>Framework</i> de desenvolvimento baseado em Java para criação e manipulação de ontologias utilizando OWL, RDF e SPARQL. Integrado na aplicação desenvolvida em Java, permitiu criar componentes para <i>match & merge</i> (integração de vários conceitos na mesma ontologia), Inferência, TDB (gravar a ontologia na base de dados Apache TDB) e para consulta à ontologia para expor o resultado em Java Swing.	Reutilização de métodos para manipulação programática de ontologias que permitiu criar o processo de <i>match & merge</i> e <i>ontology learning</i> , integrando os modelos OWL intermédios com origem em ArchiMate, Microsoft Excel e Microsoft SQLServer.
Apache OpenNLP.	<i>Framework</i> de desenvolvimento para processamento de língua natural para vários idiomas, incluindo português, que permite separação de frases em palavras, análise de sintaxe de frases e outros métodos específicos para língua natural.	Reutilização de métodos já desenvolvidos para processamento de língua natural que permitiu testar o modelo de interface com utilizador.
Java.	Linguagem de programação em <i>open source</i> que permite o desenvolvimento de vários tipos de aplicações, incluindo a possibilidade de desenvolver via Java Swing uma interface gráfica para interação com utilizadores ou desenvolvimento de aplicações <i>web</i> e <i>mobile</i> .	Linguagem base utilizada para integração com os <i>frameworks</i> ApacheJENA e OpenNLP para desenvolvimento do protótipo.

Tabela 6.4: Tecnologias utilizadas em metadados como ontologia

Na arquitetura tecnológica foram consideradas as seguintes opções:

- Foi utilizado o *plugin* ArchiOWL para conversão de ArchiMate para OWL por funcionar como *open source* para a ferramenta de modelação Archi;
- Foi testado o *plugin* DataMaster para conversão de bases de dados relacionais para OWL por funcionar integrado com o Protégé, mas tinha a limitação de só funcionar com a versão Protégé v3 e pretendia-se utilizar como base de partida uma arquitetura empresarial numa ferramenta específica como é o caso de Achi/ArchiMate;
- Face à necessidade de se tratar as tabelas, campos e restrições a partir da sua definição DDL em base de dados, optou-se por fazer uma ligação direta à base de dados Microsoft SQLServer para executar *queries* que permitem explorar os metadados DDL e programar em ApacheJENA a sua transformação para OWL;
- Foi utilizado o *framework* ApacheJENA após comparação com OWL API e Protégé OWL API, por ser mais independente e permitir no futuro outras funcionalidades eventuais por ser um projeto da Apache, onde existem bases de dados não estruturadas como o TDB que foi igualmente utilizado para armazenamento da ontologia, ou o OpenNLP para processamento de língua natural;
- Foi utilizado o *framework* OpenNLP por ter um conjunto de métodos essenciais para processamento de língua natural, conjugado com o suporte para língua portuguesa;
- Em qualquer das opções de tecnologia não foi considerada uma avaliação detalhada com as alternativas pelo facto do objetivo da investigação não ser a avaliação de tecnologias para cada funcionalidade, mas a experimentação técnica de uma hipótese de solução.

6.3.4.2 Arquitetura aplicacional

Pelo facto de ser necessário guardar ontologias em base de dados Apache TDB, preparar a ontologia carregando ficheiros OWL, DDL e Microsoft Excel, guardar configurações para cada entidade com uma ontologia, além de analisar e guardar expressões por utilizador, considerou-se um conjunto de funcionalidades a seguir descritas na Figura 6.16 ao nível de estruturas de classes Java criados e com descrições de funcionalidades na Tabela 6.5, agrupadas por níveis de funcionalidades.

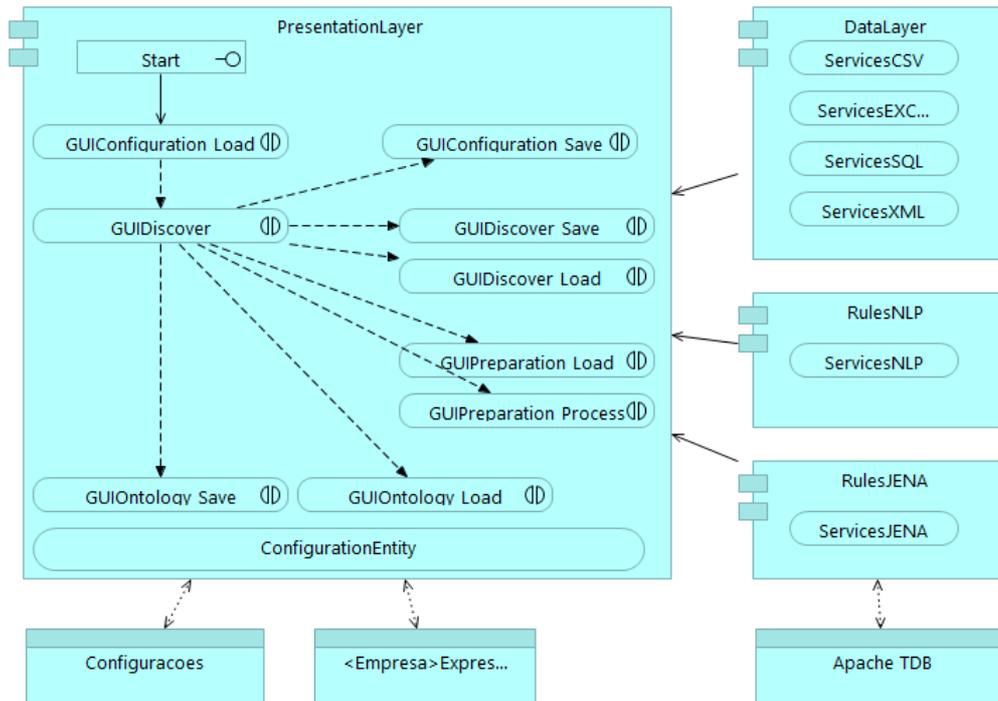


Figura 6.16: Arquitetura aplicacional metadados como ontologia

De acordo com a Figura 6.16, a solução está organizada nos seguintes *packages* java:

- **PresentationLayer:** classes com métodos específicos para criar interação com utilizador utilizando Java Swing. Todas as classes passam como parâmetro a classe *ConfigurationEntity* que tem o contexto da navegação com informação da entidade em tratamento e expressão em análise;
- **DataLayer:** classes para cada tipo de dados (Microsoft Excel, CSV, XML e Microsoft SQLServer) com métodos específicos para ler e escrever;
- **RulesJENA:** classe com métodos específicos para criar modelos, ler modelos diretamente, gerar inferência, ler modelos via SPARQL, gravar modelos em Apache TDB e ler modelos de Apache TDB;
- **RulesNLP:** classe com métodos específicos para separar frases num texto (*SentenceDetector*), separar palavras numa frase (*tokenizer*) e análise sintática e semântica de palavras (*parte-of-speech*, *parsing*).

Além da base de dados Apache TDB com o modelo de ontologia gravado, a aplicação utiliza ainda 2 tipos de ficheiros de texto:

- **Configuracoes:** mantém informação sobre cada entidade, descrição, data, servidor, base de dados, *user/password* de acesso a base de dados e ficheiro de expressões a utilizar onde são guardadas as expressões por opção do utilizador;

- **<empresa>Expressoes:** mantém informação sobre cada expressão que o utilizador queira guardar para reutilização, com dados sobre expressão, data, dono e descrição. Empresa é substituído com a entidade em causa.

Nas classes acima referidas, são implementados vários métodos específicos para tratamento de dados tal como apresentados na Tabela 6.5:

Função	Comportamento	Motivação
Acesso a dados.	Conjunto de classes com métodos específicos para acesso a dados em XML, CSV, Microsoft Excel e Microsoft SQLServer.	Isolar métodos que podem ser reutilizados.
Regras Jena.	Conjunto de métodos que permitem ler ficheiros OWL para o modelo, carregar o modelo OWL a partir de CSV, carregar o modelo a partir de DDL, gerar inferência e pesquisar o modelo OWL, além de métodos para gravar e ler em base de dados Apache TDB.	Isolar métodos de manipulação de estruturas OWL e regras de negócio.
Regras NLP.	Conjunto de métodos que permitem separar frases (<i>SentenceDetector</i>), separar palavras (<i>Tokenizer</i>) e analisar sintaticamente as palavras na frase (<i>Parte-of-speech</i>).	Isolar métodos de manipulação de estruturas NLP e regras de negócio.

Tabela 6.5: Classes Java de acesso a dados

No caso da camada de apresentação para exploração de dados utilizando Java Swing, foram implementadas as classes apresentadas na Tabela 6.6:

Função	Comportamento	Motivação
Guardar e ler ontologia (GUIOntology_save/load).	Gravar e ler em base de dados Apache TDB a ontologia OWL criada, associado metadados específicos da mesma.	Permitir reutilização à posteriori.
Guardar e ler configurações (GUIConfiguration_save/load).	Gravar em ficheiro de configuração por entidade, a informação sobre a entidade ao nível de servidor de base de dados de <i>business intelligence</i> e ficheiro a utilizar para guardar expressões analisadas.	Permitir que a experimentação seja multiempresa.
Integrar os vários modelos que compõem a ontologia (GUIPreparation_load/process).	Permitir ler ficheiro OWL, Microsoft Excel, DDL, com modelos complementares de conceitos para a ontologia. Com base na ontologia criada, permitir gerar classificações de métricas e dimensões utilizando inferência.	Separar o processo de carregamento de conceitos do processamento (utilizando inferência).
Analisar e guardar expressões (GUIDiscover_save/load).	Permitir analisar expressões, guardar e recuperar expressões, por utilizador.	Permitir reutilizar as mesmas expressões pelo utilizador.

Tabela 6.6: Classes Java de exploração de dados

No caso de gravar e ler ontologia, apresentado na Figura 6.17, identifica-se primeiro a entidade em causa, e no caso gravar, detalha-se as suas características.

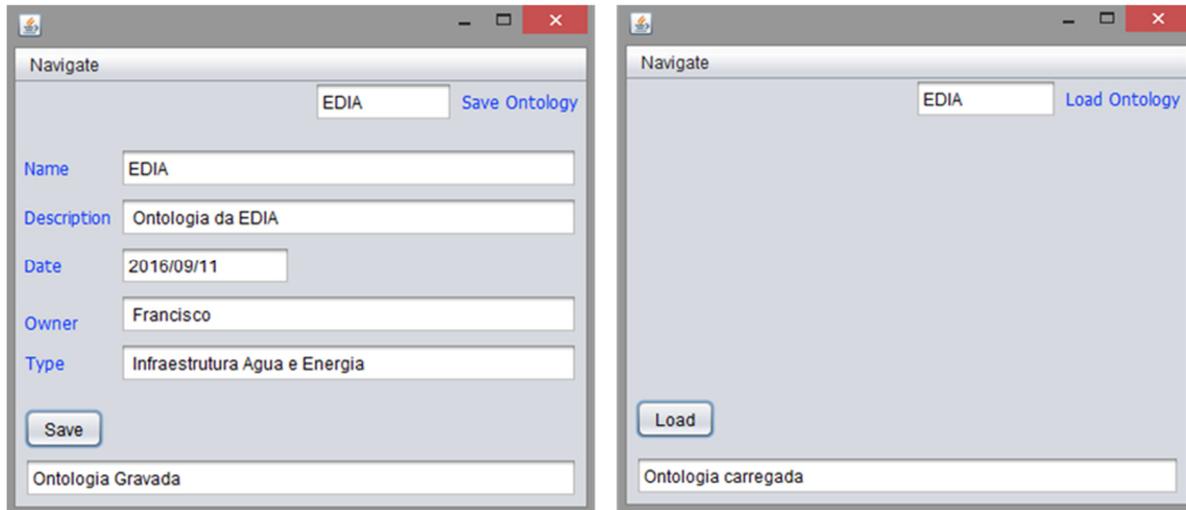


Figura 6.17:Gravar e ler ontologia

No caso de gravar e ler configurações da entidade, apresentado na Figura 6.18, seleciona-se a entidade em causa no caso de ler configurações, ou detalha-se as configurações para a entidade específica.

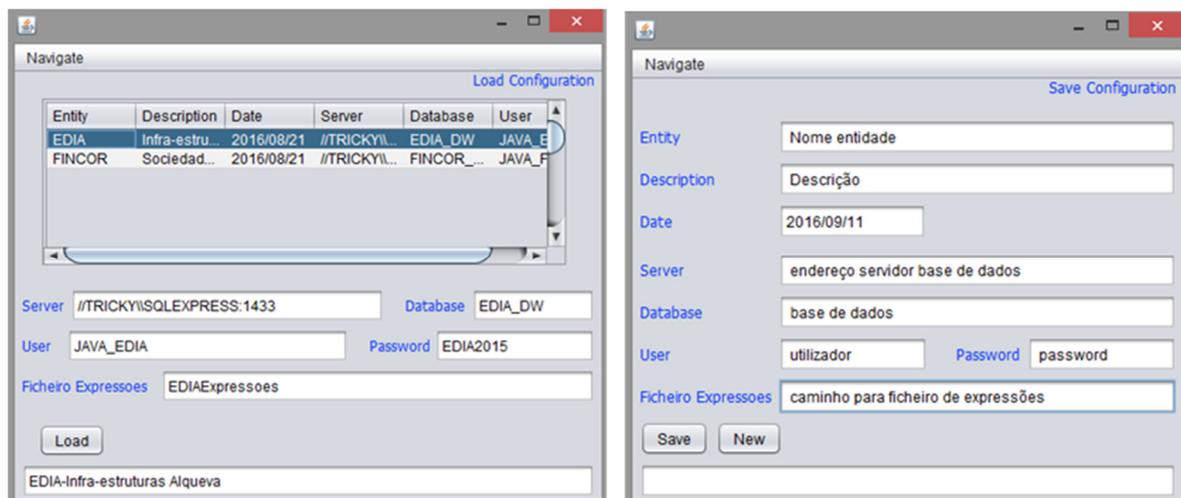


Figura 6.18: Ler e guardar configurações

Para se integrar e processar a ontologia, tal como apresentado na Figura 6.19, indica-se o tipo de conceitos a integrar, sendo que se deve utilizar a sequência “Arquitetura Empresarial ArchiMate”, “DDL” e “Glossário”, para se carregar a ontologia OWL. De seguida, pode-se efetuar o processamento, que por via de inferência gera as classificações em termos de dimensões, métricas, domínios de informação e itens de glossário.

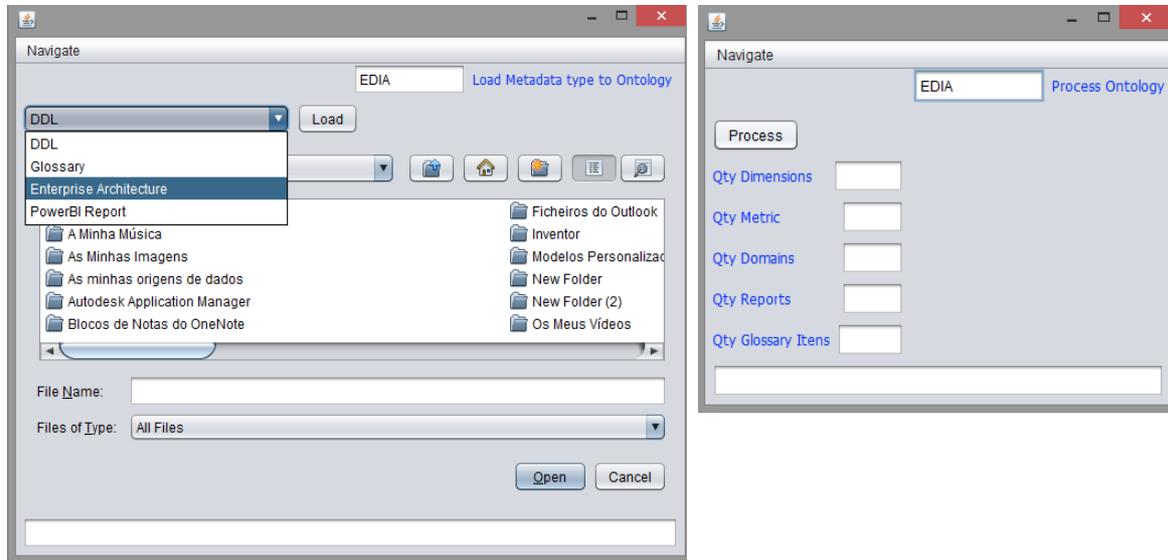


Figura 6.19: Interpretar e processar ontologia

Para se analisar as expressões, tal como apresentado na Figura 6.20, indica-se a expressão, a partir do qual se aplica o NLP *Sentence Detector* para separar o caso de ser indicada mais do que uma frase, separada por “.”, seguido de NLP *Tokenizer* e NLP *Parte-Of-Speech*, para obter e classificar as palavras na primeira fase, a partir da qual se utiliza SPARQL para se identificar o que são métricas, dimensões, domínios e restrições. Para cada conceito é apresentada a sua estrutura (conceito, tipo BI, conceito AE, campo e tabela base de dados) e ao selecionar um conceito é apresentada a semântica (relação, tipo relação, com quem se relaciona e classe do destino da relação). Como resultado, mostra-se a estrutura e semântica de cada conceito na expressão, além de se mostrar a estrutura de SQL DML associado à expressão mas sem instanciar tabelas e campos, mantendo somente os conceitos numa estrutura SQL. As expressões podem ser gravadas, para posterior utilização, tal como apresentado na Figura 6.21.

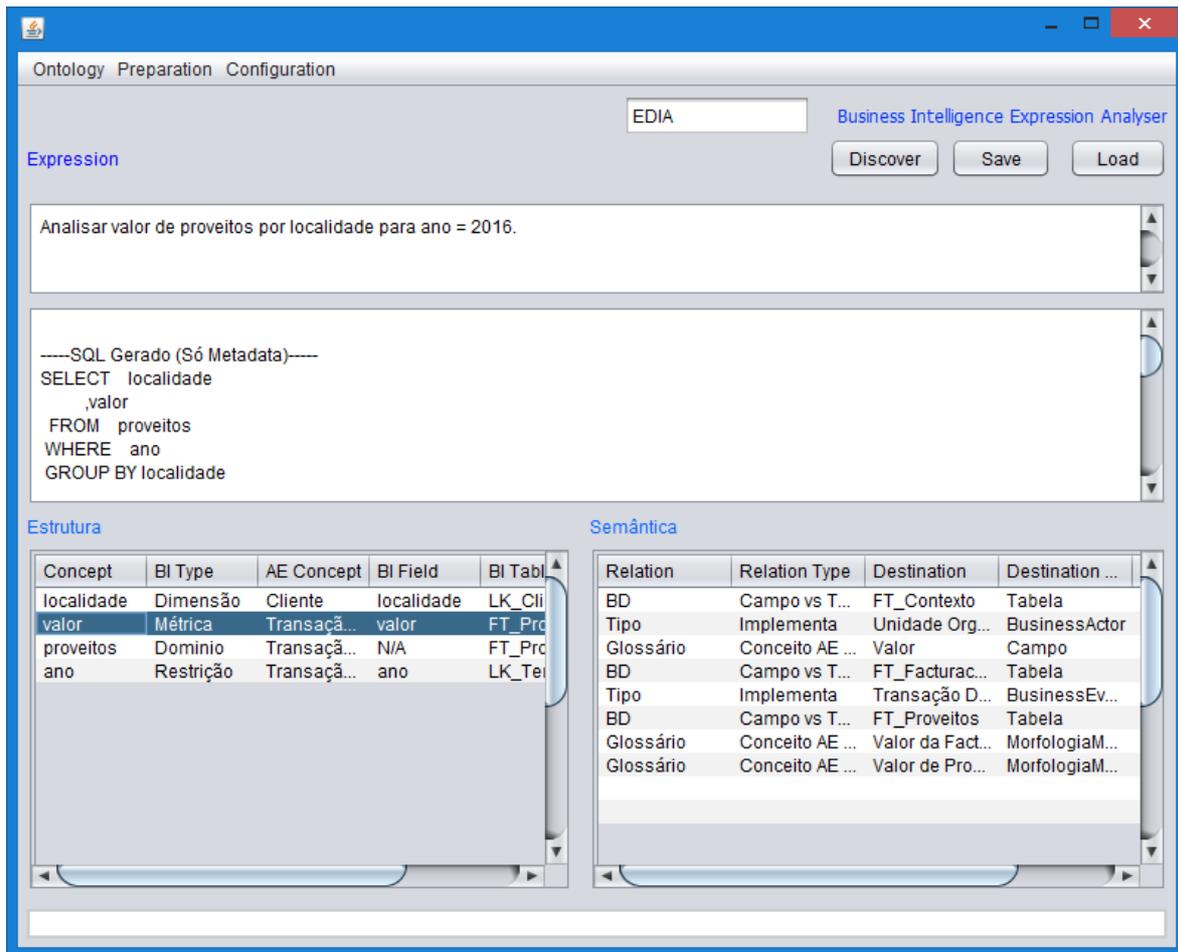


Figura 6.20: Análise de expressões de informação de gestão

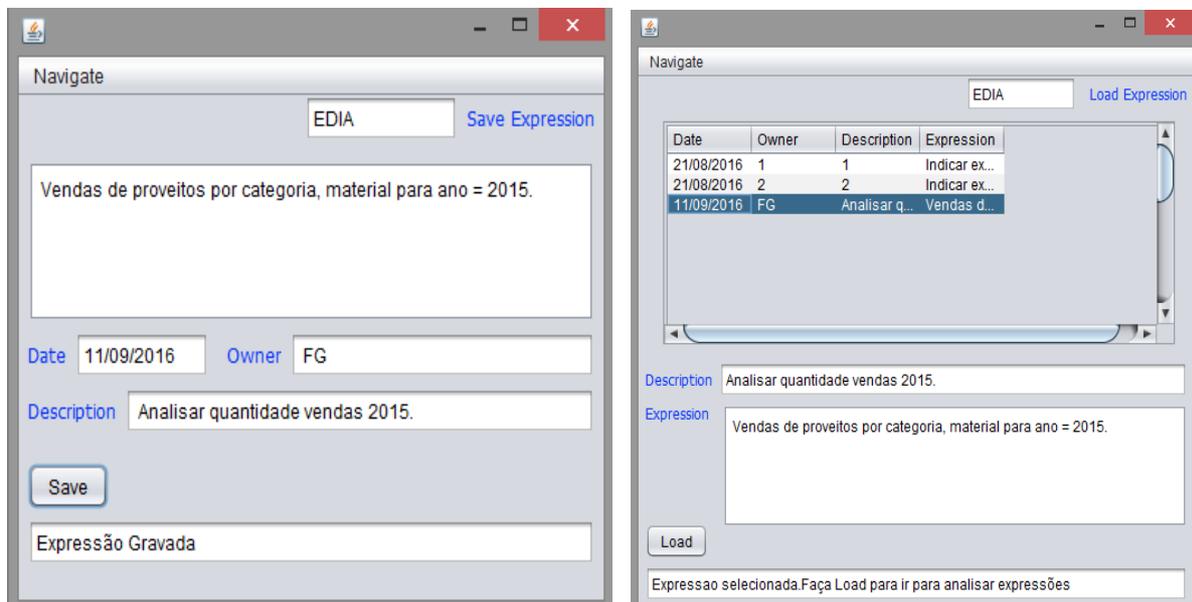


Figura 6.21: Gravar e ler expressões de informação de gestão

6.3.4.3 Integração de ontologia

Para carregamento da ontologia a partir de AE, DDL e Glossário, criaram-se em Java várias classes de acordo com o modelo 6.16 com métodos específicos para o efeito, isolando cada tipo de modelo a integrar numa única ontologia, seguindo a sequência AE, seguida de DDL e por fim Glossário. Esta sequência deve-se ao facto de que primeiro é necessário criar os conceitos chave definidos em AE, para de seguida se poder seguir um modelo de integração de várias ontologias numa só.

No caso da integração do modelo AE na ontologia, é processado a partir do método `GUI_Preparation_Load`, que evoca o método `ServicesJenaImportarArchimate`. Este método por sua vez evoca o método `OWL_CriarClassesBase` que cria de forma programada um conjunto de classes OWL consideradas como âncora no protótipo (classes gerais que são independentes da organização e de seguida evoca o método `OWL_ImportarArchimate` para importação do ficheiro OWL (com os conceitos específicos da organização). A especificação destes métodos é a seguinte:

- No caso do método `OWL_CriarClassesBase`, cria a ontologia, as classes base (e.g. `ArquiteturaEmpresarial`, `BaseDados`, `Tabela`, `Campo`, `Formato`) e propriedades (e.g. `caracteriza`, `contem`, `SignificaAE`) por detalhe do modelo de metadados indicado na Figura 6.2;
- No caso do método `OWL_ImportarArchimate`, acede ao ficheiro gerado pela ferramenta Archi para importar para OWL via método `ApacheJENA OntModel.read(ficheiroOWL)` para uma ontologia criada pelo método anterior.

Para integração do DDL é necessário carregar as tabelas e campos, mas também as chaves estrangeiras. Como tal, considera-se o seguinte:

- O método `GUI_Preparation_Load` evoca o método `ServicesJenaImportarDDL` que por sua vez evoca o método `OWL_ImportarDDL` para importar as tabelas e campos tendo como informação o resultado da *query* DML indicada na Figura 6.22 que retorna por cada campo, a base de dados, tabela e formato associado, sendo que os campos são carregados como indivíduos da classe “Campo” e as tabelas na classe “Tabela”;
- O método `GU_Preparation_Load` evoca o método `ServicesJenaImportarDDL` que por sua vez evoca o método `OWL_ImportarDDLChavesEstrangeiras` para importar para a ontologia as chaves estrangeiras em termos de tabela e campo origem, e tabela e campo destino como indivíduos da ontologia. Para o efeito este método utiliza a *query* DML indicada na Figura 6.23 que retorna informação por restrição com informação para se integrar na ontologia.

```
String xSelectSQL = "SELECT T.name AS Tabela "
                  +",T.modify_date AS TabelaDataModificacao "
                  +",T.create_date AS TabelaDataCriacao "
                  +",C.name AS CampoNome "
                  +",F.name AS CampoFormato "
                  +",C.max_length AS CampoTamanho "
                  +",C.precision AS CampoCasasDecimais "
                  +",C.is_identity AS CampoChave "
                  +",C.is_nullable AS CampoNulo "
                  +",K.CONSTRAINT_NAME "
                  +",Z.type AS TipoObjecto "
                  +",Z.type_desc as DescObjecto "
                  +" FROM sys.columns C "
                  +" JOIN sys.tables T "
                  +"   ON c.object_id = t.object_id "
                  +" JOIN sys.types F "
                  +"   ON c.user_type_id = F.user_type_id "
                  +" LEFT OUTER JOIN INFORMATION_SCHEMA.KEY_COLUMN_USAGE AS K "
                  +"   ON K.TABLE_NAME = T.name "
                  +"   AND K.COLUMN_NAME = C.name "
                  +" LEFT OUTER JOIN sys.objects Z "
                  +"   ON z.name = k.CONSTRAINT_NAME ";
```

Figura 6.22: Query DML para extrair metadados SQLServer (Tabelas e Campos)

```
String xSelectSQL = "SELECT FKColunas.constraint_object_id as BaseID "
                  +",Base.name as BaseNome "
                  +",FKColunas.parent_object_id as OrigemID "
                  +",OrigemTabela.name as OrigemTabela "
                  +",OrigemColuna.name as OrigemCampo "
                  +",FKColunas.referenced_object_id as DestinoID "
                  +",DestinoTabela.name as DestinoTabela "
                  +",DestinoColuna.name as DestinoCampo "
                  +" FROM sys.foreign_key_columns FKColunas "
                  +" LEFT OUTER JOIN sys.objects Base "
                  +"   ON Base.object_id = FKColunas.constraint_object_id "
                  +" LEFT OUTER JOIN sys.objects OrigemTabela "
                  +"   ON OrigemTabela.object_id = FKColunas.parent_object_id "
                  +" LEFT OUTER JOIN sys.columns OrigemColuna "
                  +"   ON OrigemColuna.object_id = FKColunas.parent_object_id "
                  +"   AND OrigemColuna.column_id = FKColunas.parent_column_id "
                  +" LEFT OUTER JOIN sys.objects DestinoTabela "
                  +"   ON DestinoTabela.object_id = FKColunas.referenced_object_id "
                  +" LEFT OUTER JOIN sys.columns DestinoColuna "
                  +"   ON DestinoColuna.object_id = FKColunas.referenced_object_id "
                  +"   AND DestinoColuna.column_id = FKColunas.referenced_column_id ";
```

Figura 6.23: Query DML para extrair metadados SQLServer (Chaves Estrangeiras)

Por fim, para integração do glossário, o método GUI_Preparation_Load evoca o método ServicesJenImportarGlossário que por sua vez evoca o método OWL_ImportarGlossario. Para carregar a ontologia utiliza-se dados em Microsoft Excel com a estrutura indicada na Figura 6.24 que tem informação para cada conceito sobre a origem (tipo e elemento) e destino (significado e descrição). Os tipos podem ser “campo”, “tabela” ou qualquer elemento de ArchiMate como “BusinessActor”, “BusinessObject”, entre outros. Com base nesta informação cria-se cada individuo na classe “Glossário” e nas classes de Morfologia, além das relações com os campos, tabelas e conceitos AE.

Tipo	Elemento	Significado	Descricao
BusinessActor	Administração	Conselho Administração	Conselho Administração
Campo	Ambito	Âmbito	Âmbito
Campo	Ano	Ano	Ano
Campo	Ano	Ano	Ano
Campo	Ano	Ano	Ano
Campo	Ano	Ano	Ano
Campo	AnoMes	Ano e Mês	Ano e Mês
Campo	AnoOrig	Ano Origem	Ano Origem
Campo	AreaM2	Área em Metros Quadrados	Área em Metros Quadrados

Figura 6.24: Exemplo de Glossário

6.3.4.4 Análise de expressões de negócio

No domínio de *business intelligence* utilizam-se expressões de negócio, expressas de forma interrogativa aplicada à informação da organização.

Para análise destas expressões, consideramos o processamento de linguagem natural apresentado na Figura 6.25, onde a partir de uma expressão num texto se efetua uma decomposição em frases. Para cada frase analisa-se a sintaxe e semântica, para permitir identificar cada palavra e posicionar o seu entendimento com base na ontologia, para permitir descrever a partir da expressão o que são dimensões, métricas, domínios, restrições e sua correspondência com conceitos de metadados técnicos.

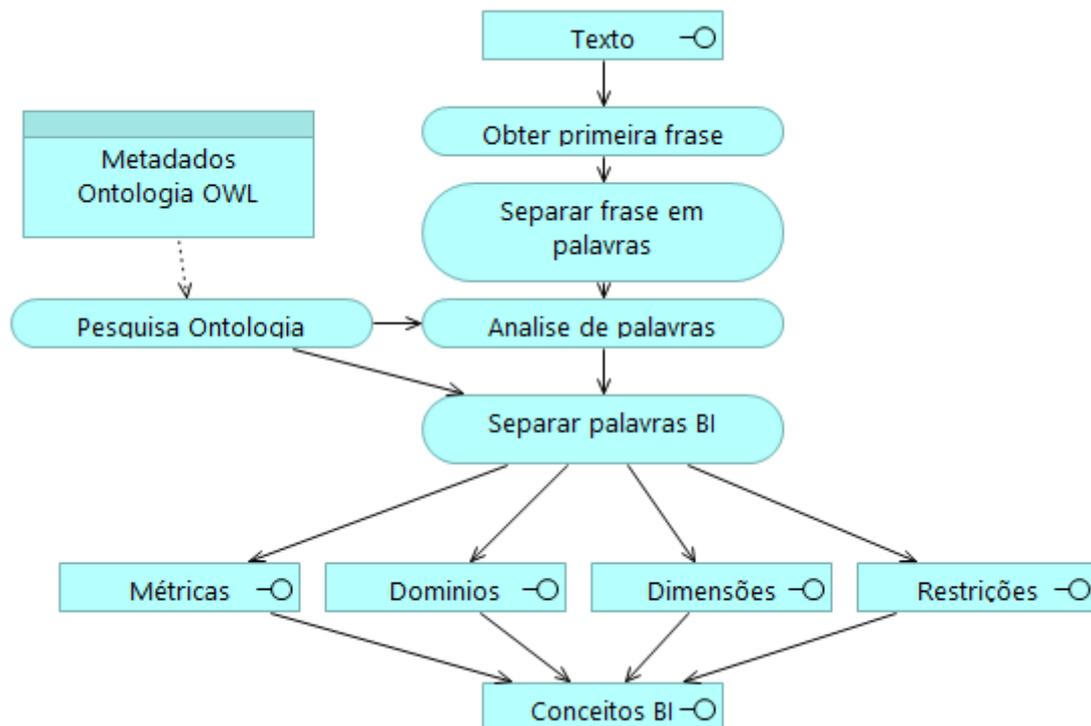


Figura 6.25: Análise de expressões de negócio

Em cada fase, são aplicadas as seguintes regras, descritas na Tabela 6.6.

Fase	Descrição	Considerações
Obter primeira fase.	Aplica-se o método <i>SentenceDetector</i> do OpenNLP para separar um texto em frases, separadas por “.”.	Utiliza-se o modelo pt-sent.bin do idioma Português disponibilizado pelo <i>framework</i> .
Separar frase em palavras.	Utiliza-se o método <i>Tokenizer</i> do OpenNLP para obter uma lista de palavras separadas por espaços.	Utiliza-se o modelo pt-token.bin do idioma Português disponibilizado pelo <i>framework</i> .
Análise de palavras.	Utiliza-se o método <i>Parte-of-Speech</i> para análise das palavras na frase.	Utiliza-se o modelo pt-pos-maxent.bin do idioma Português disponibilizado pelo <i>framework</i> .
Separar palavras BI.	Utilizando-se as palavras sem considerar números, artigos, pronomes, preposições, conjunções e advérbios, parte-se do seu enquadramento na frase (DE, POR, PARA) e obtém-se o seu enquadramento na ontologia.	Implementação de algoritmo próprio. A “Pesquisa Ontologia” é implementada via ApacheJENA com métodos para pesquisa de indivíduos e navegação por classes e propriedades da ontologia.

Tabela 6.7: Análise gramatical de expressão

6.4 Resultados esperados

Tendo por base a arquitetura descrita no sub-capítulo 6.3 face ao problema apresentado no capítulo 4 e avaliado junto a empresas no capítulo 5, tenciona-se obter os seguintes resultados na experimentação em casos de estudo, alinhado com as questões de investigação definidas no sub-capítulo 1.2:

- Avaliar a adequação de alinhamento entre conceitos definidos na arquitetura empresarial e em *business intelligence*, considerando metadados como ontologia e o papel crítico de dimensões e métricas neste alinhamento;
- Avaliar formas de classificar as métricas e dimensões com base num modelo de ontologia derivado da arquitetura empresarial;
- Utilizar a ontologia como forma de comunicação com utilizadores para descoberta de conhecimento através de linguagem natural;
- Demonstrar o papel crítico da arquitetura empresarial nas organizações enquanto forma de representação da ontologia organizacional, mas devidamente alinhado com *business intelligence* onde é efetuada a monitorização de desempenho organizacional, em torno do mesmo modelo da organização;
- Utilizar formas de visualizações de informação em *business intelligence*, como é o caso de *dashboard*, para explorar informação organizada da mesma forma que é consultada numa arquitetura empresarial, mas com vantagens ao nível de instanciação de valores e navegação mais facilitada entre a informação.

A solução em termo de modelo e tecnologia utilizada abre igualmente linhas de investigação futuras como descrito no sub-capítulo 9.5.

7 Casos de estudo

Neste capítulo apresentam-se a experimentação da hipótese de arquitetura de solução aplicada a dois casos de estudo. É apresentado o caso de uma empresa de gestão de infraestrutura de águas e energia (sub-capítulo 7.2) e o de uma empresa de corretagem em bolsa (sub-capítulo 7.3). No final, é apresentado o questionário consolidado que foi respondido pelas empresas (sub-capítulo 7.4).

7.1 Introdução

Os casos de estudo foram utilizados para permitirem uma avaliação da relevância do problema e para se testar a hipótese de solução em casos reais. Para o efeito, consideraram-se duas organizações:

- **Organização no sector de infraestrutura de água e energia:** empresa que produz água e energia, sendo que o produto final relevante é a água distribuída numa região delimitada e tendo como objetivo a irrigação de campos agrícolas, apesar de poder ser comercializada para consumo em casos de urgência;
- **Organização no sector financeiro:** empresa de corretagem financeira que trabalha com clientes particulares e empresas em operações de compra e venda de títulos financeiros em bolsa em qualquer parte do mundo recorrendo a *brokers* internacionais.

Para cada caso de estudo, analisou-se a organização para identificar o tipo de modelação que efetuam e que utilização fazem de sistemas de *business intelligence*. Com base na informação recolhida foi realizado o seguinte:

- Desenhou-se o modelo de arquitetura empresarial em hipótese, ao nível da componente de “Organização” e “Negócio”, utilizando a ferramenta Archi e ajustando o modelo ao caso específico. Para cada componente criou-se uma estrutura de ontologia sem elementos, e outra instanciada com elementos, para permitir comparar com a visualização via *dashboard*;
- Analisaram-se os dados operacionais relativos à informação sobre “Organização/Custos” e “Negócio/Proveitos”, para de seguida se montar um modelo de dados em *Microsoft SQLServer* alimentado a partir dos dados operacionais, utilizando-se *Microsoft Integration Services* e criando um modelo de exploração dos dados via *dashboard* com *Microsoft PowerBI Desktop*, tal como definido na hipótese de solução;
- Com base nos glossários de conceitos de cada organização, modelo de dados criado, relatórios e modelos representados na arquitetura empresarial, carregou-se a ontologia específica para cada caso, utilizando-se o protótipo desenvolvido. Uma vez carregada a ontologia, analisaram-se expressões de negócio e pesquisou-se termos relevantes face à ontologia, vista desta forma como metadados corporativos.

Pelo facto de serem empresas reais, foram retiradas as métricas de valores financeiros, centrando-se somente em métricas de quantidade e mesmo essas, com valores alterados. Por outro lado, foram retirados dados de ambientes de teste e mascarando nomes e valores, por questões de confidencialidade de dados das organizações.

7.2 Empresa gestora de Infraestruturas de água e energia

É uma entidade gestora de uma infraestrutura de barragens e canais para produção de energia e captura de água para regadio, além de serviços complementares de turismo de natureza e museologia.

A base do negócio é a captura de água e a produção de energia elétrica, suportada em infraestruturas primárias (e.g. barragens, perímetros, blocos) para captura/produção, e, infraestruturas secundárias (canais, hidrantes, contadores) para distribuição da água até ao consumidor.

A organização distribui diretamente a água aos consumidores, sendo que estes podem ser particulares ou empresas. No caso da energia esta está concessionada a outra entidade. Como negócio complementar, a organização tem serviços de cartografia, turismo com Parque de Natureza e museologia, entre outros. Como resumo da sua oferta, esta organização disponibiliza produtos (água e serviços), comercializados a clientes em segmentos (particulares, empresas, governo) através de canais presenciais.

Para efeito da investigação, considerou-se um subconjunto de componentes do sistema de informação para análise:

- **Sistema operacional:** 26 sistemas aplicativos estruturados entre ERP, gestão de infraestruturas físicas, construção e comercial. Os sistemas estão em várias tecnologias, implementados por vários fornecedores com várias bases de dados e sem uma visão integrada de dados;
- **Sistema *business intelligence*:** sistema de indicadores com recolha manual e *dashboard*. São feitas várias *queries* para pesquisa e utilizados mecanismos de pesquisa do ERP SAP. Como tal, construiu-se um sistema de raiz, utilizando-se tecnologia Microsoft BI, para permitir testar-se a arquitetura proposta como hipótese;
- **Sistemas de representação de conhecimento:** não é utilizado um sistema de arquitetura empresarial. Os sistemas de informação e o modelo da organização é caracterizado em documentação em formato Microsoft Word e Microsoft Excel no contexto de políticas, processos e procedimentos.

7.2.1 Arquitetura empresarial

7.2.1.1 Arquitetura organizacional

Ao nível da ontologia, sem instanciação de elementos tal como apresentada na Figura 7.1, pode-se destacar a necessidade de criar um grupo com a estrutura orgânica de acordo com a relação entre os tipos de unidades (e.g. administração, direção, departamento), mas depois teve-se que criar uma generalização do conceito para servir de base para relacionar com funções que cada unidade orgânica necessita e os contratos FSE (fornecimentos e serviços externos) que estabelecem.

Considerando a estrutura e semântica específica do negócio desta organização, os contratos são estabelecidos e geridos por unidades orgânicas, envolvendo fornecedores e imobilizado (e.g. edifícios, carros, software, hardware), sendo que esses contratos geram transações de custos. Por outro lado, o imobilizado está sediado em determinados locais, sendo que nesses locais existem igualmente colaboradores. Os colaboradores assumem funções em unidades orgânicas e estabelecem contratos de trabalho com a organização. Estes contratos geram transações de custos com pessoal. Temos assim que as métricas estão associadas a estas transações e os restantes conceitos correspondem a dimensões.

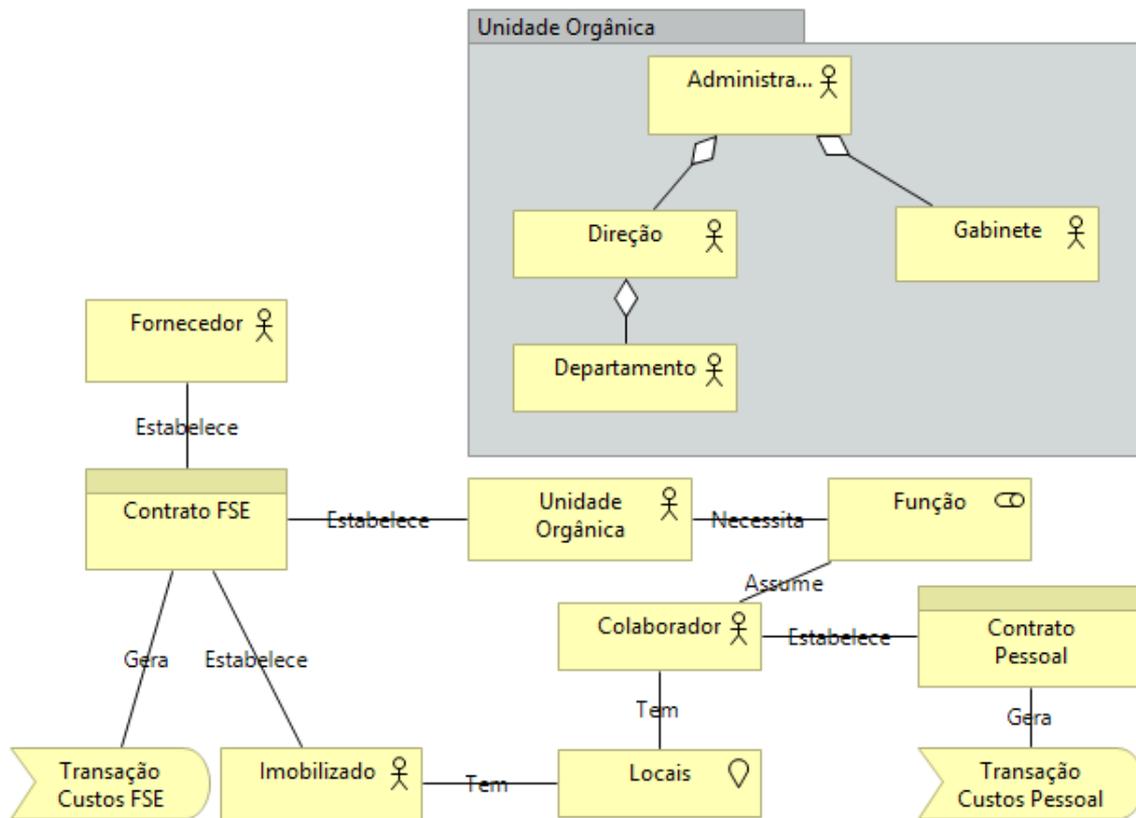


Figura 7.1: Modelo de ontologia da arquitetura organizacional de empresa gestora de infraestruturas

A ontologia instanciada (detalhada) numa lógica de arquitetura empresarial é apresentado na Figura 7.2, para mais tarde se comparar com a visualização via *dashboard*, mas sem ser exaustiva em casos como fornecedores, pessoal por função, face à quantidade de elementos que teriam que ser instanciados. Sendo essa uma das vantagens de se utilizar *dashboard* em *business intelligence*.

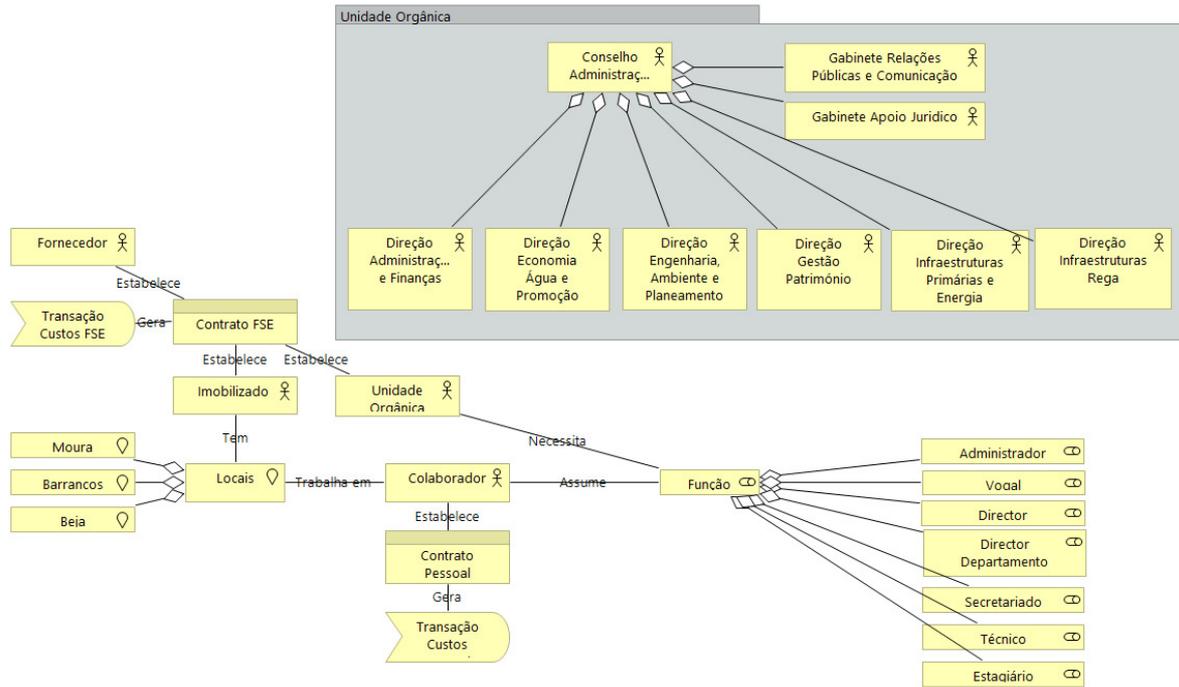


Figura 7.2: Modelo instanciado de ontologia da arquitetura organizacional de empresa gestora de infraestruturas

O modelo de arquitetura de informação correspondente à arquitetura organizacional é apresentado na Figura 7.3. Neste modelo destaca-se a necessidade de relacionar as estruturas de dados em termos de orgânica (LK_EstruturaOrganica) e centros de gestão (LK_CentroGestao) com o mesmo conceito em AE. O caso inverso acontece com os conceitos “Contrato FSE” e “Contrato Pessoal” que estão ambos associados a uma única tabela de dados (FT_Custos). Por outro lado, os conceitos que não têm tabelas de dados associados devem-se somente ao facto de se ter delimitado o âmbito da modelação, pois em princípio cada conceito corresponde a uma tabela ou a um campo de dados, sendo que deveria garantir-se na medida do possível que um conceito correspondesse univocamente a uma tabela.

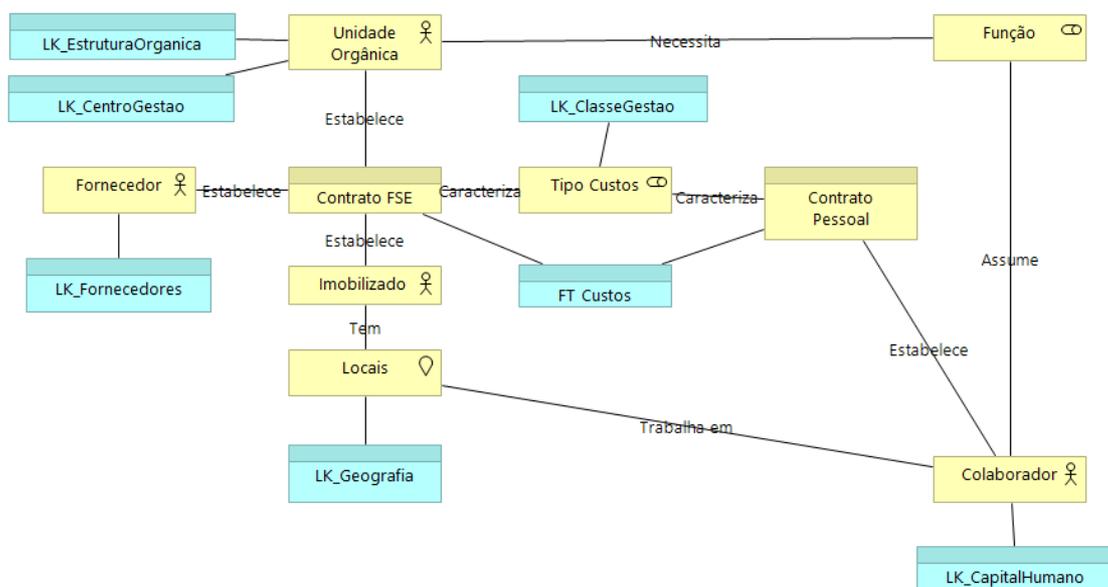


Figura 7.3: Modelo de relação entre arquitetura organizacional e arquitetura de informação de empresa gestora de infraestrutura

7.2.1.2 Arquitetura negócio

Ao nível da ontologia, sem instanciação de elementos, tal como apresentada na Figura 7.4, a entidade gere uma infraestrutura composta por sistemas, perímetros e blocos para produção de eletricidade além de armazenamento de água. A organização gere igualmente museus temáticos.

Não considerando o negócio de produção de eletricidade, esta organização utiliza canais para distribuir a oferta comercial através de serviços que são objeto de contratos de proveitos estabelecidos com clientes ou por transação direta em casos em que não são necessários contratos (e.g. visitas aos museus, serviços de cartografia). As transações estão assim associadas a contratos ou relacionam a oferta com os clientes. Os clientes são tipificados por segmentos (e.g. particular, empresa, instituições financeiras, instituições governamentais).

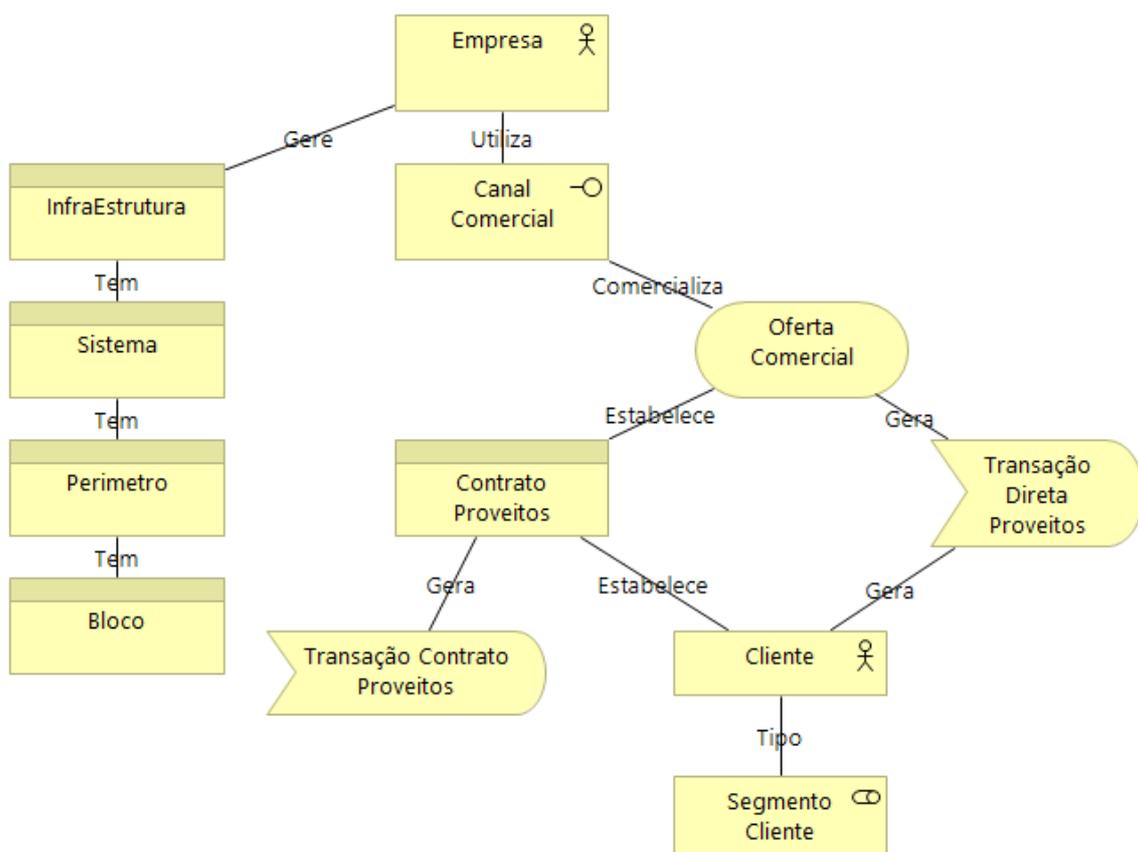


Figura 7.4: Modelo de ontologia da arquitetura negócio de empresa gestora de infraestruturas

O modelo instanciado numa lógica de arquitetura empresarial é apresentado na Figura 7.5, para mais tarde se comparar com a visualização via *dashboard*. Não são representados neste caso as instâncias de elementos de infraestrutura (e.g. bloco em concreto), clientes, contratos e transações, face à quantidade em causa, sendo essa uma das vantagens de se consultar essa informação via *dashboard em business intelligence*.

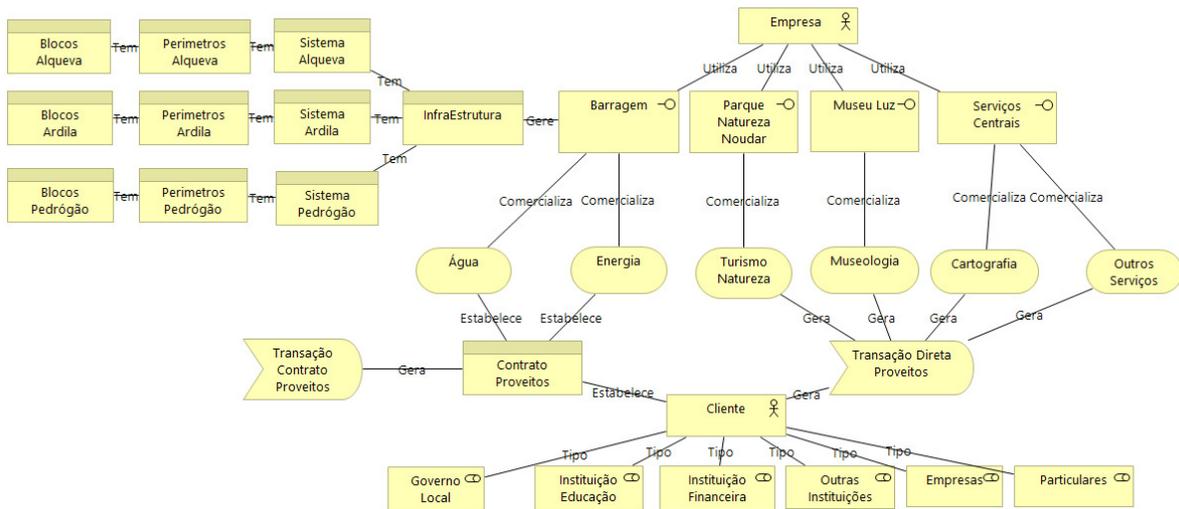


Figura 7.5: Modelo instanciado de ontologia da arquitetura negócio de empresa gestora de infraestruturas

O modelo de arquitetura de informação correspondente à arquitetura de negócio é apresentado na Figura 7.6.

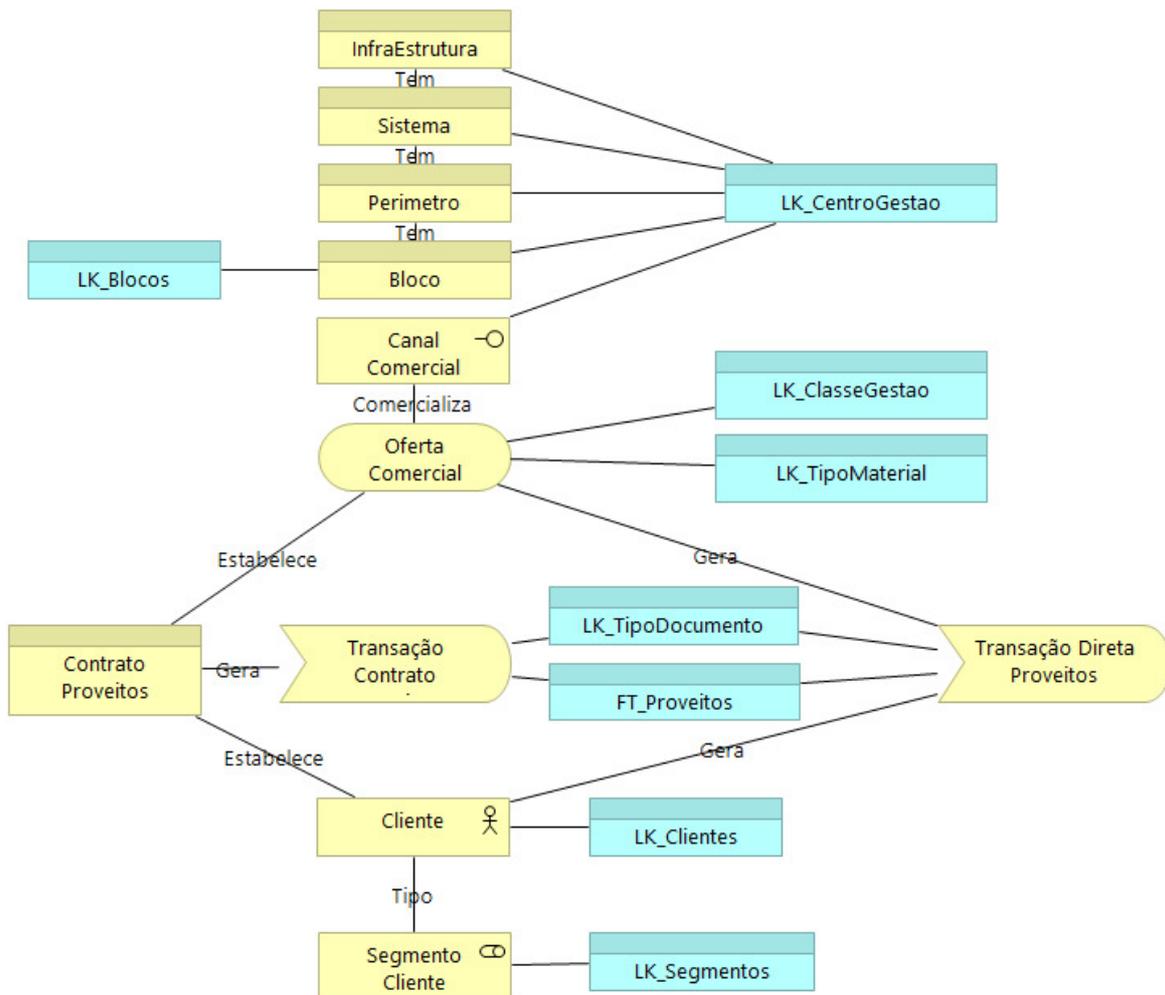


Figura 7.6: Modelo de relação entre arquitetura de negócio e arquitetura de informação de empresa gestora de infraestrutura

7.2.2 Business intelligence

7.2.2.1 Base de dados

O modelo apresentado na Figura 7.7 e relações entre tabelas apresentadas na Figura 7.8 foi criado em Microsoft PowerBI Desktop a partir de tabelas em Microsoft SQLServer. A estrutura da base de dados foi desenhada tendo por base os conceitos da organização face à definição em arquitetura empresarial sendo assim uma implementação da ontologia sobre esta forma de representação de conhecimento. O modelo é caracterizado pelo seguinte:

- Tabela FT_Proveitos como elemento central do modelo em *snowflake* para proveitos a partir do qual se estruturam as várias tabelas de dimensões, como é o caso da dimensão de clientes (LK_Clientes) a segmentação de clientes (LK_Segmento), ou o caso do tipo de documento de faturação (LK_TipoDocumento) ou tipo de material (LK_TipoMaterial);
- Tabela FT_Custos como elemento central do modelo *snowflake* para custos a partir do qual se se estruturam tabelas de dimensões como é o caso de fornecedores (LK_Fornecedores);
- Tabelas LK_CentroGestao e LK_ClasseGestao que descrevem a infraestrutura e unidades orgânicas, ligadas a estruturas de custos (FT_Custos) e proveitos (FT_Proveitos).

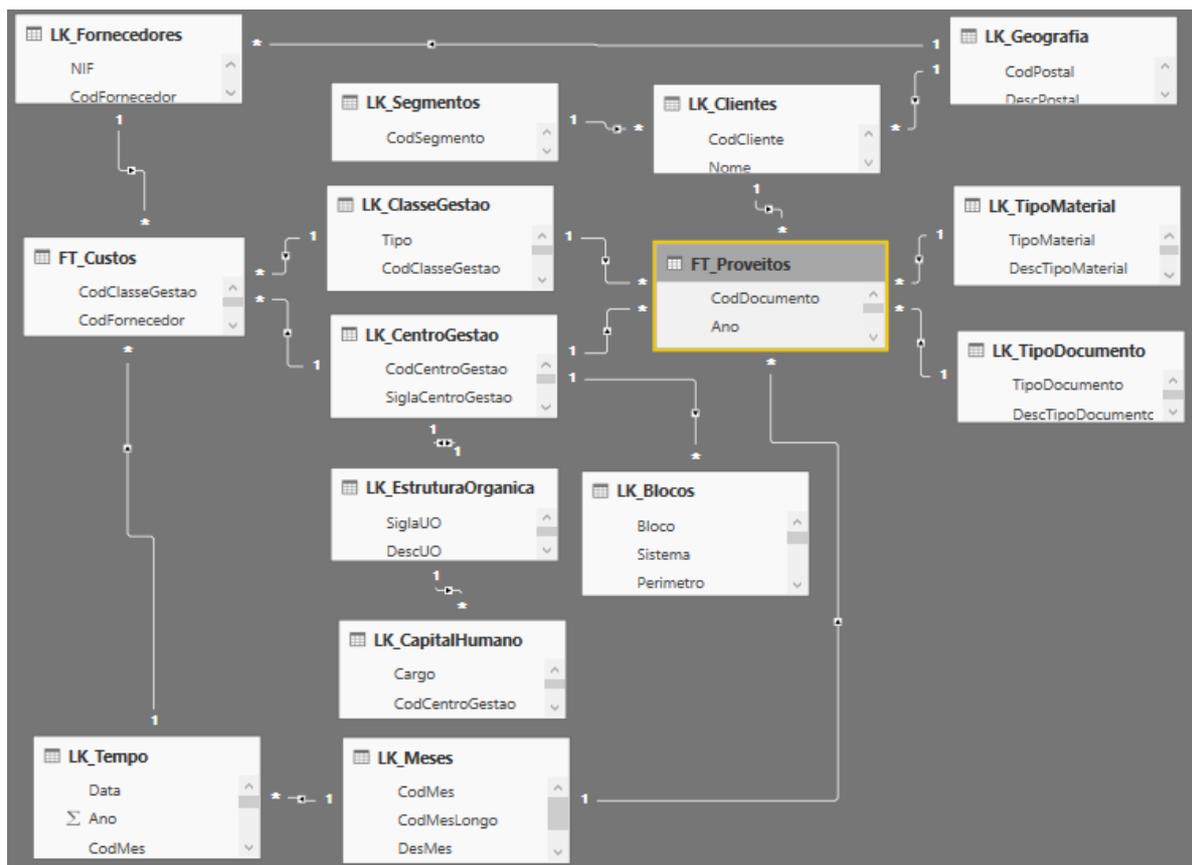


Figura 7.7: Modelo de dados empresa gestora de infraestruturas em PowerBI Desktop

Manage relationships

Active	From: Table (Column)	To: Table (Column)
<input checked="" type="checkbox"/>	FT_Custos (CodCentroGestao)	LK_CentroGestao (CodCentroGestao)
<input checked="" type="checkbox"/>	FT_Custos (CodClasseGestao)	LK_ClasseGestao (CodClasseGestao)
<input checked="" type="checkbox"/>	FT_Custos (CodFornecedor)	LK_Fornecedores (CodFornecedor)
<input checked="" type="checkbox"/>	FT_Custos (DataDocumento)	LK_Tempo (Data)
<input checked="" type="checkbox"/>	FT_Proveitos (CodCentroGestao)	LK_CentroGestao (CodCentroGestao)
<input checked="" type="checkbox"/>	FT_Proveitos (CodClassGestao)	LK_ClasseGestao (CodClasseGestao)
<input checked="" type="checkbox"/>	FT_Proveitos (CodCliente)	LK_Clientes (CodCliente)
<input checked="" type="checkbox"/>	FT_Proveitos (Periodo)	LK_Meses (CodMesLongo)
<input checked="" type="checkbox"/>	FT_Proveitos (TipoDocumento)	LK_TipoDocumento (TipoDocumento)
<input checked="" type="checkbox"/>	FT_Proveitos (TipoMaterial)	LK_TipoMaterial (TipoMaterial)
<input checked="" type="checkbox"/>	LK_Blocos (CodCentroGestao)	LK_CentroGestao (CodCentroGestao)
<input checked="" type="checkbox"/>	LK_CapitalHumano (CodCentroGestao)	LK_EstruturaOrganica (CodCentroGestao)

New... Autodetect... Edit... Delete

Close

Figura 7.8: Modelo de relação entre tabelas da empresa gestora de infraestruturas

Como resultado, o modelo deste caso de estudo é caracterizado pelo seguinte:

- Tabela central de factos FT_Proveitos associada a dimensões críticas de clientes (LK_Clientes), transações (LK_TipoDocumento e LK_TipoMaterial), produtos (LK_ClasseGestao) e responsabilidades pelos proveitos (LK_CentroGestão);
- Tabela central de factos FT_Custos associada a fornecedores (LK_Fornecedores), tipo de compras (LK_ClasseGestão) e responsabilidade pela compra (LK_CentroGestao);
- Tabela de clientes (LK_Clientes) como dimensão, mas com uma relação com segmentação (LK_Segmento) e localizações (LK_Geografia);
- Tabela de fornecedores (LK_Fornecedores) como dimensão, mas com uma relação com localizações (LK_Geografia);
- Tipo de vendas (LK_TipoMaterial e LK_TipoDocumento) com forma de identificar o tipo de transação;
- Produtos e serviços comercializados sob a forma de Classe de Gestão (LK_ClasseGestao);
- Responsabilidade sobre os custos e proveitos por um conceito de Centro de Gestão (LK_CentroGestao), reutilizado igualmente para elementos da estrutura orgânica;
- Estrutura orgânica (LK_Estrutura Orgânica) e sua relação com colaboradores (LK_CapitalHumano);
- Utilização de tempo (LK_Tempo) como forma de introduzir o conceito de histórico no modelo, como é normal em modelos informacionais. Neste caso foi colocado só ao nível de tabelas de factos, apesar da modelação poder incluir também esta dimensão em tabelas de dimensões.

7.2.2.2 Integração de dados

O processo de ETL (*extraction, transform and load*) foi criado utilizando-se a ferramenta Microsoft *Integration Services*. Os processos foram organizados em agrupamentos com a mesma lógica de conceitos da arquitetura empresarial tal como apresentado na Figura 7.9. Por essa razão, foram criados dois grupos, “Orgânica” e “Negócio” onde se consideraram tarefas específicas de carregamento por tabela. As tabelas transversais como LK_Geografia, LK_Tempo, LK_CentroGestao e LK_ClasseGestao foram colocadas fora destes agrupamentos.

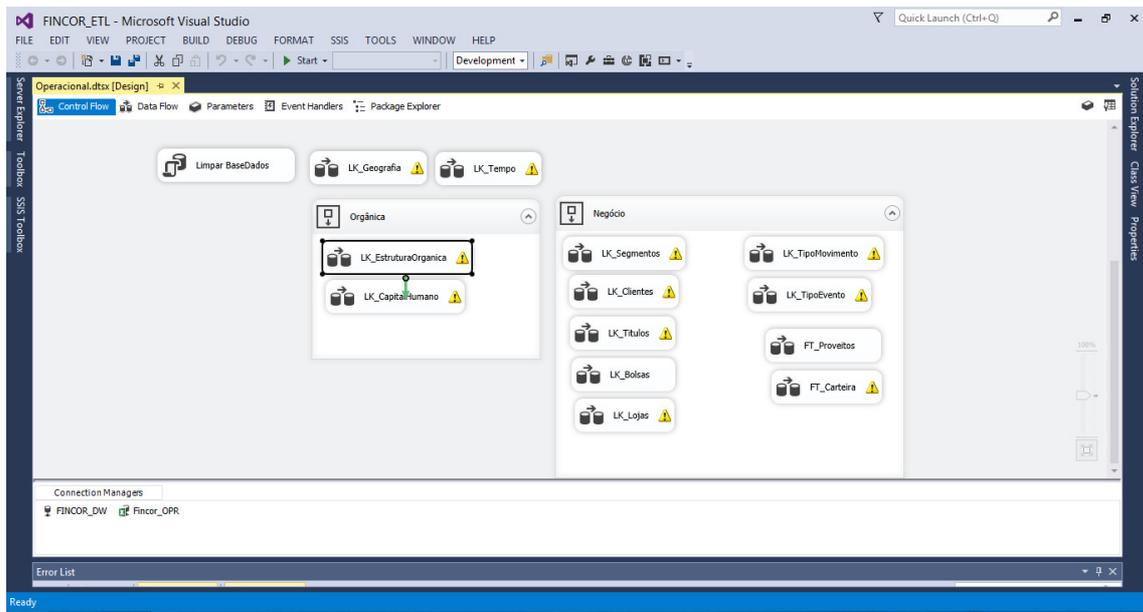


Figura 7.9: Processo ETL empresa gestora de infraestruturas

Por outro lado, tal como apresentado na Figura 7.10, para cada tarefa existe um processo de ligação a tabelas origem (*extraction*), seguido de uma transformação (*transform*) e por fim o carregamento na tabela destino (*load*). Neste caso não foram efetuadas muitas transformações pois o modelo de base já tinha os campos preparados para carregamento.

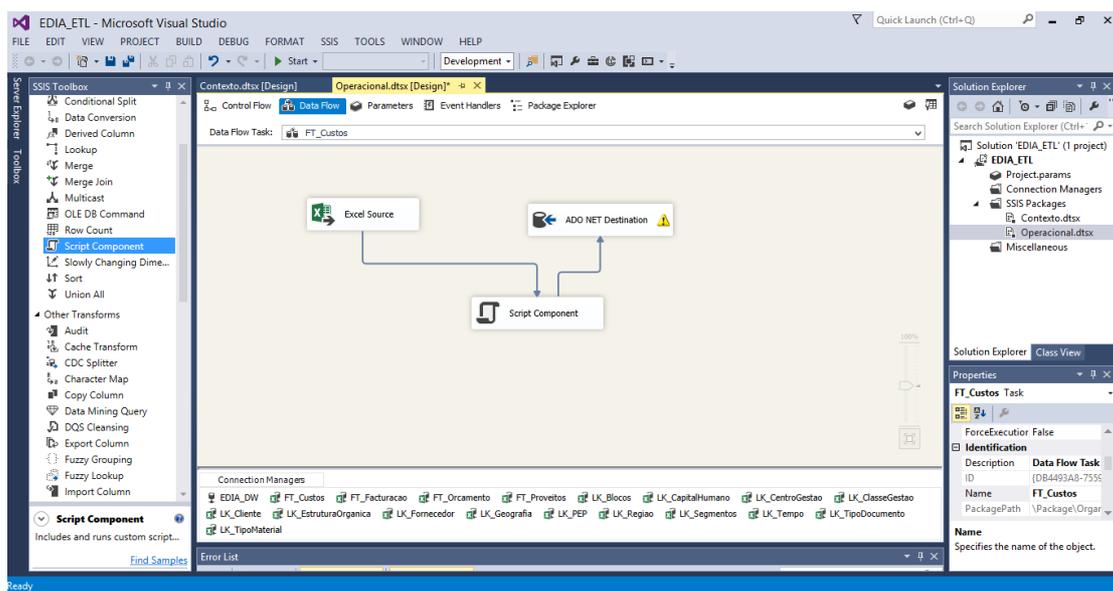


Figura 7.10: Processo ETL empresa gestora de infraestruturas. Exemplo de transformação por tabela

7.2.2.3 Dashboard para exploração de dados

Face à orientação da hipótese de solução, foi criado um documento com quatro folhas (*sheets de acordo com o Microsoft Power BI*) relativas a “Estrutura organizacional”, “Estrutura negócio”, “Custos” e “Proveitos” enquanto áreas temáticas de análise da organização. De notar que os dados apresentados são fictícios visto terem sido mascarados a partir de dados da empresa em caso de estudo.

Na *sheet “estrutura organizacional”* na Figura 7.11, pode-se observar a estrutura orgânica montada com base num gráfico “heat map” para permitir “drill” a partir da hierarquia orgânica sendo que a dimensão das zonas do gráfico representa a quantidade de colaboradores. Por outro lado, face aos locais de atividade da empresa, é apresentado sobre a forma de um mapa as zonas do país com o tamanho das “bolas” relacionadas com a quantidade de colaboradores. São ainda apresentadas dimensões críticas de análise de colaboradores, como género, nacionalidade e cargo (equivalente ao conceito “função” na ontologia, mas era um termo assim reconhecido na organização pelo que se fez este ajuste na designação). São assim apresentados os conceitos de estrutura orgânica, colaboradores e locais, que fazem parte da modelação em arquitetura empresarial, agora instanciados em elementos concretos, utilizando-se as quantidades como métricas e toda a estrutura de conceitos representados numa arquitetura empresarial, como se fossem atributos para detalhar as análises até ao detalhe de cada elemento (e.g. pessoa, unidade orgânica, localidade).

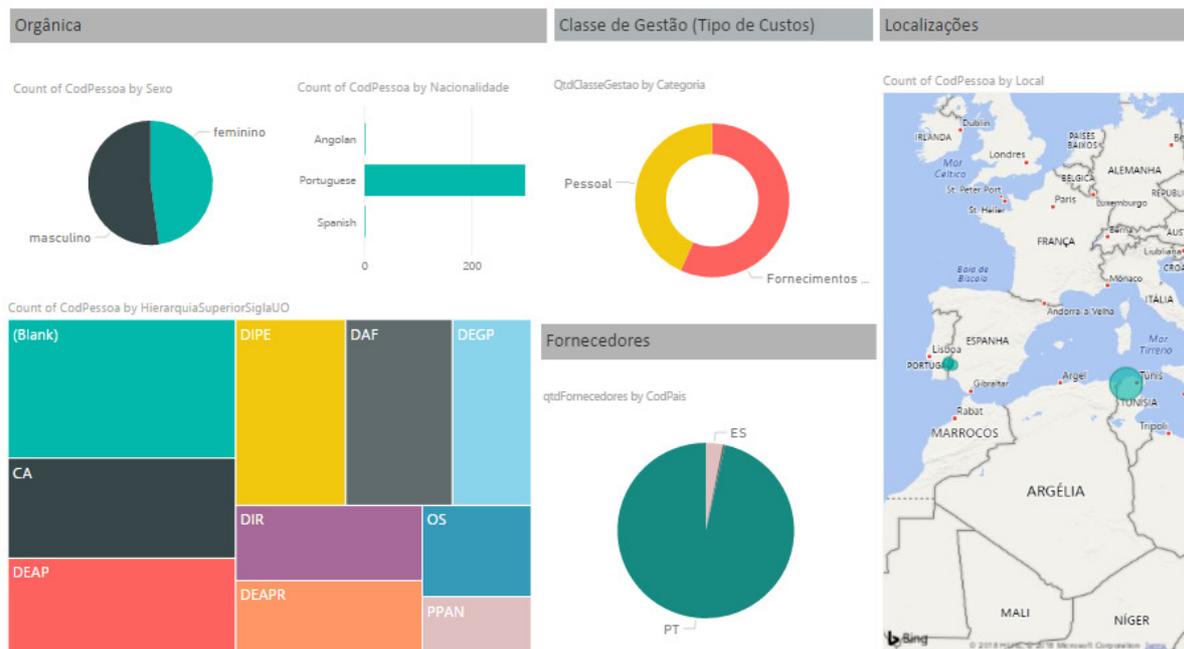


Figura 7.11: Sheet “Estrutura Organizacional” do dashboard de empresa gestora de infraestruturas

Na *sheet “estrutura negócio”* na Figura 7.12, pode-se observar a representação em gráfico para navegação de quantidades de componentes de infraestrutura de acordo com a hierarquia de sistema, perímetro e bloco. Por outro lado, são apresentados igualmente os tipos de custos, responsabilidades de custos, oferta comercial, fornecedores e segmentação.

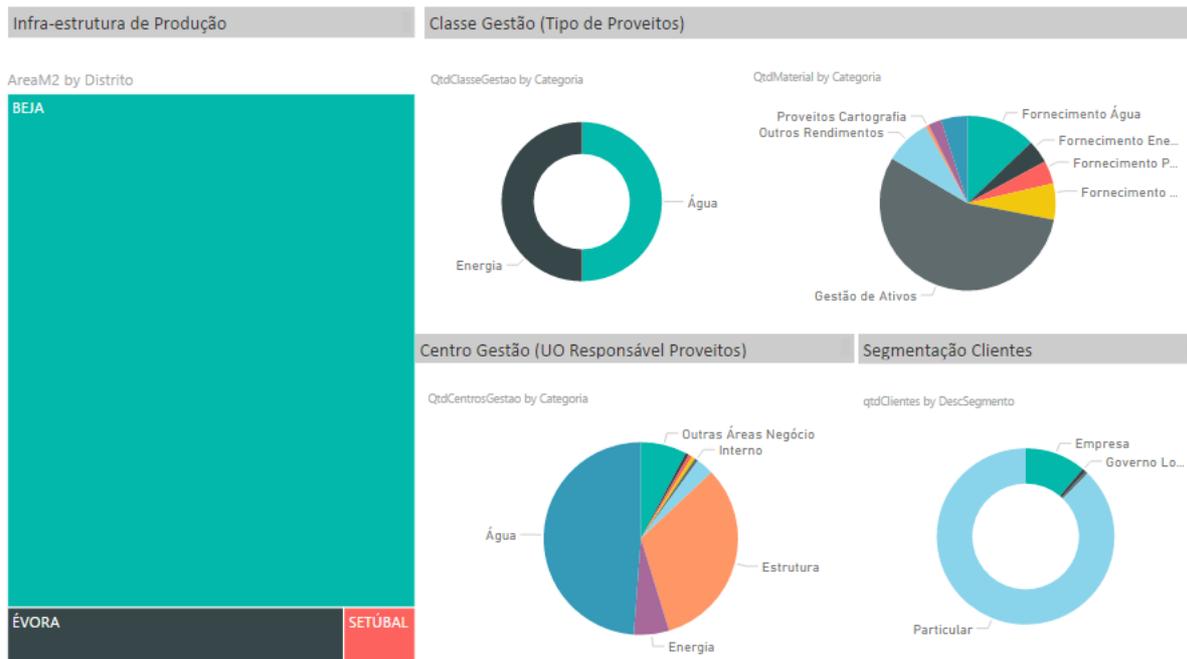


Figura 7.12: Sheet “Negócio” do dashboard de empresa gestora de infraestruturas

Na sheet “proveitos” na Figura 7.13, pode-se observar que são reutilizadas as dimensões ao nível de “arquitetura de negócio” relativas a oferta comercial e segmentação de clientes, mas acrescentado somente uma visão de tempo associado às transações e quantidade de transações.

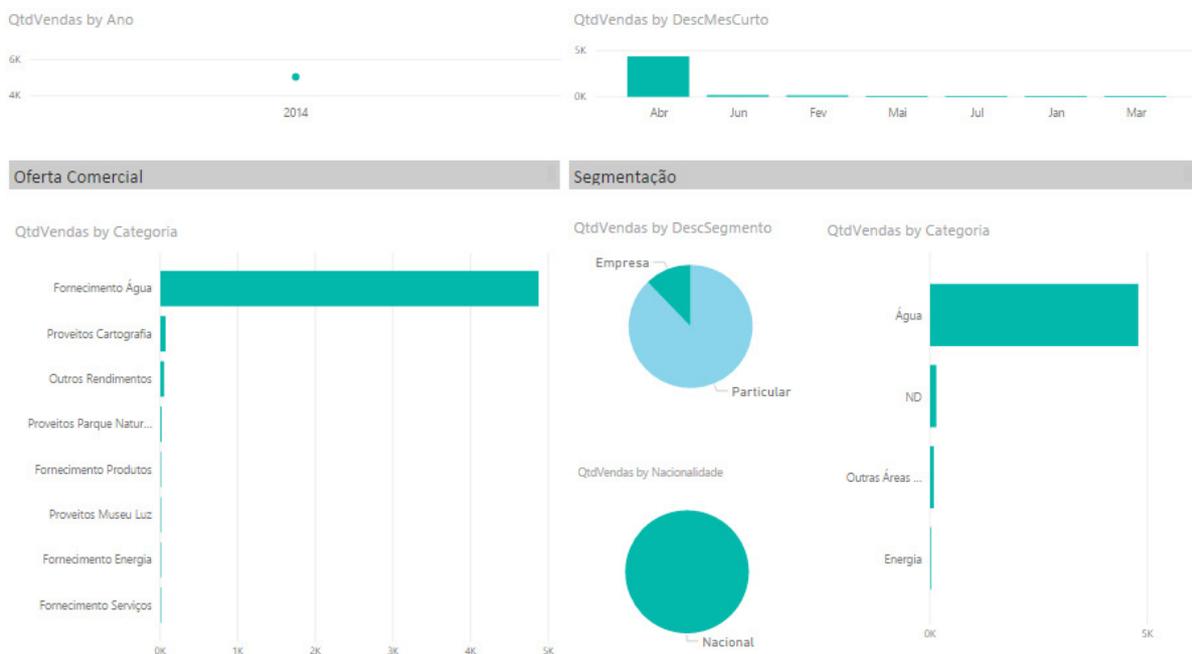


Figura 7.13: Sheet “Proveitos” do dashboard de empresa gestora de infraestruturas

Na *sheet* “custos” na Figura 7.14, pode-se observar que são reutilizadas as dimensões ao nível de “arquitetura de negócio” relativas a tipo de custo, responsabilidade dos custos (nesta organização este conceito está associado ao que designam por centro de gestão) e fornecedores, mas acrescentando somente uma visão de tempo associado às transações e quantidade de transações.



Figura 7.14: Sheet “Custos” do *dashboard* de empresa gestora de infraestruturas

De notar na visualização de *dashboard* que as expressões “métricas por dimensões” (isto é, analisa lista de métricas por dimensões) é nativa na própria ferramenta, como se repara na legenda que a ferramenta coloca em gráficos (e.g. qtdcompras por categoria; qtdcompras por CodPais).

7.2.3 Analisador de expressões de negócio

Testando algumas expressões com a experimentação de Processamento de Língua Natural reutilizando a ontologia criada, temos um resultado adequado, tal como apresentado na Figura 7.15 para uma expressão do tipo “analisar valor DE proveitos POR localidade PARA ano=2016”:

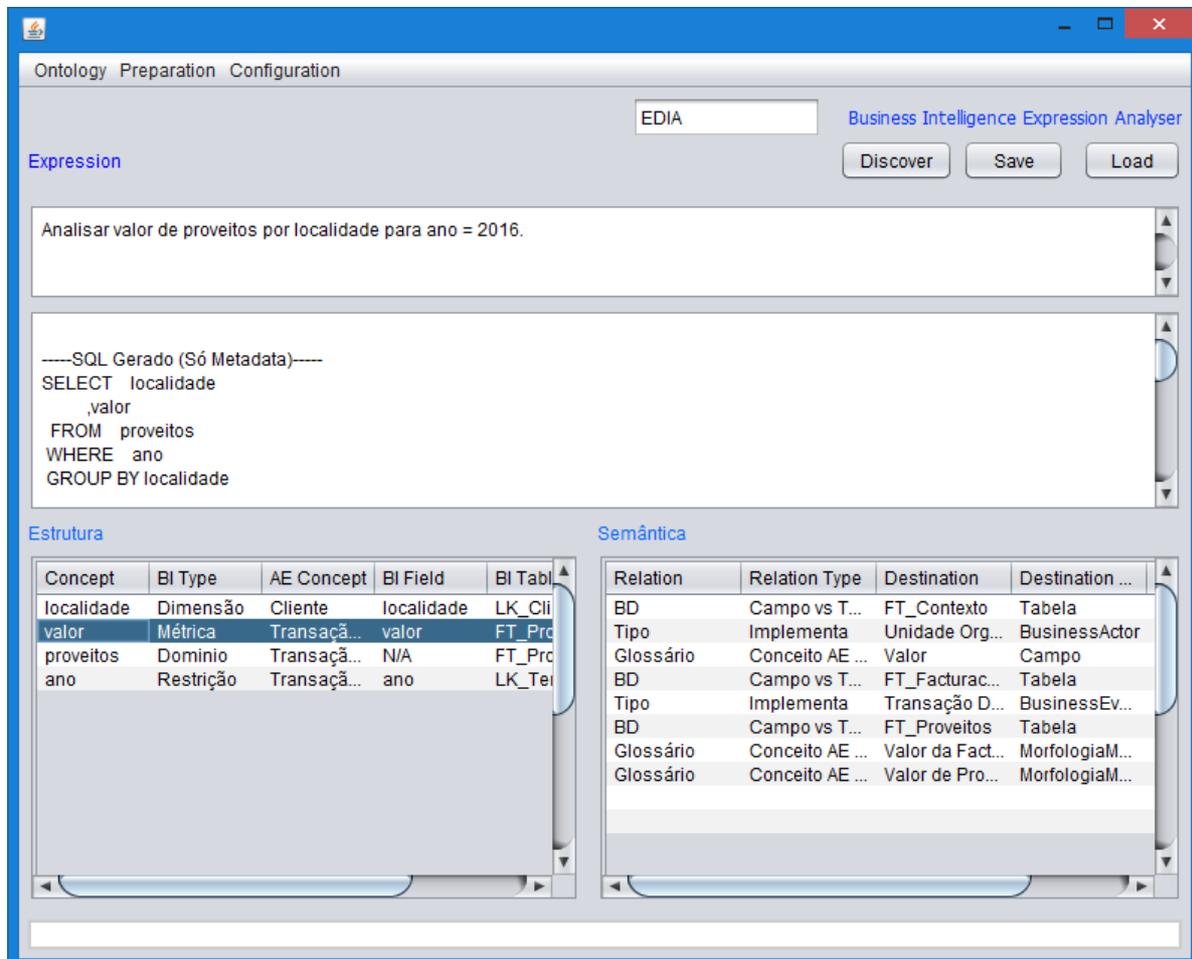


Figura 7.15: Analisador de expressões de negócio de empresa gestora de infraestruturas

7.3 Empresa financeira de corretagem

A Instituição Financeira do caso de estudo é uma corretora financeira que efetua operações de compra e venda de instrumentos financeiros (e.g. acções, obrigações, unidades de participação) por conta de solicitações de clientes, não tendo como tal carteira própria para negociação de títulos.

A base do negócio assenta na segmentação de clientes (empresas, particulares), produtos (e.g. acções, obrigações e fundos de investimento), comercializados por serviços de corretagem (*trading* e custódia de títulos) através de canais presenciais (parceiros e lojas) e dependente de parceiros (e.g. bolsas e *brokers*).

Para a investigação considerou-se um subconjunto de componentes do sistema de informação para análise:

- **Sistema operacional:** tem um sistema ERP para contabilidade, compras e fornecedores e um sistema de processamento de negócio com um *front-office* e *back-office*, além de sistemas complementares para relação com clientes no canal não presencial internet;
- **Sistema *business intelligence*:** não tem um sistema de *business intelligence*, sendo que se extrai vários ficheiros do sistema de processamento de negócio para análise de informação de gestão em Microsoft Excel;
- **Sistema de representação de conhecimento:** não é utilizado um sistema de arquitetura empresarial. Os sistemas de informação e o modelo da organização é caracterizado em documentação em formato Microsoft Word e Microsoft Excel no contexto de políticas, processos e procedimentos.

7.3.1 Arquitetura empresarial

7.3.1.1 Arquitetura organizacional

Ao nível da ontologia, sem elementos instanciados, tal como apresentada na Figura 7.16, pode-se destacar a necessidade de criar um grupo com a estrutura orgânica de acordo com a relação entre os tipos de unidades (e.g. administração, direção, departamento), mas depois teve-se que criar uma generalização do conceito para servir de base para relacionar com funções que cada unidade orgânica necessita e os contratos FSE (fornecimentos e serviços externos) que estabelecem.

Os contratos de FSE (fornecimentos e serviços externos) são estabelecidos e geridos por unidades orgânicas, envolvendo fornecedores e imobilizado (e.g. edifícios, carros, software, hardware), sendo que esses contratos geram transações de custos. Por outro lado, o imobilizado está sediado em determinados locais, sendo que nesses locais existem igualmente colaboradores. Os colaboradores assumem funções em unidades orgânicas e têm contratos de trabalho com a organização (“Contratos Pessoal”). Estes contratos geram transações de custos com pessoal. Temos assim que as métricas estão associadas às transações de custos, sendo que os restantes conceitos correspondem a dimensões.

O modelo é equivalente ao apresentado no caso de estudo anterior pois as arquiteturas organizacionais dependem do mesmo tipo de conceitos pois são genéricos, razão pela qual na arquitetura de solução tentou-se generalizar os mesmos para depois se testar esta realidade em cada caso de estudo. Os ajustamentos resultam normalmente de campos de informação específicos a recolher em cada organização.

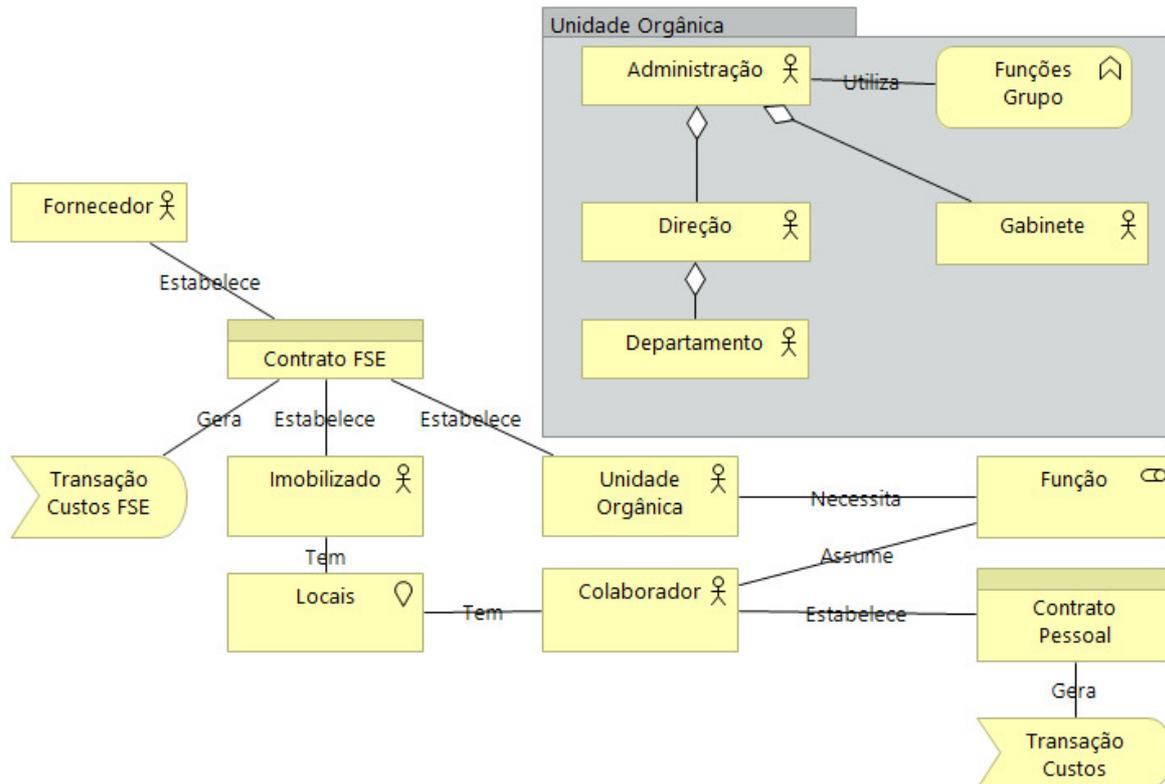


Figura 7.16: Modelo de ontologia da arquitetura organizacional de empresa financeira

O modelo instanciado numa lógica de arquitetura empresarial é apresentado na Figura 7.17, para mais tardar se comparar com a visualização via *dashboard*. Não são representados elementos de contratos, imobilizado, fornecedores e pessoal por função, pelo impacto visual complexo que tal acarretava, sendo essa uma das vantagens de se utilizar *dashboard* em *business intelligence*.

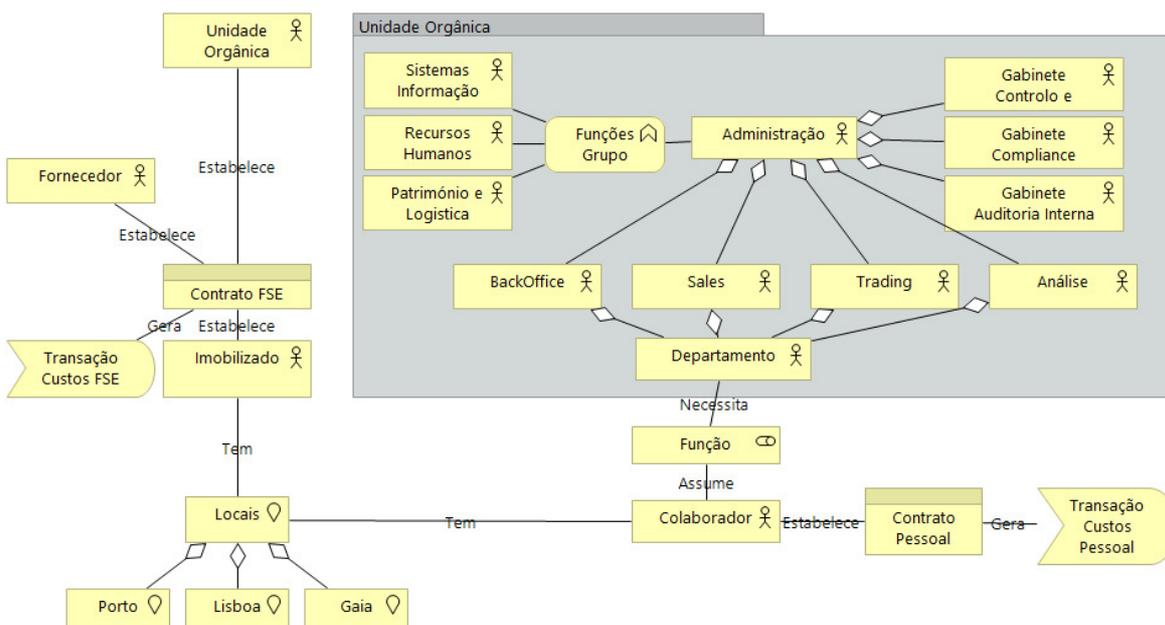


Figura 7.17: Modelo instanciado de ontologia da arquitetura organizacional de empresa financeira

O modelo de arquitetura de informação correspondente à arquitetura organizacional é apresentado na Figura 7.18, onde se destaca um alinhamento mais direto entre conceito e tabela. A exceção são conceitos como “função”, “contrato FSE” e “contrato pessoal” onde não foi possível modelar e testar os dados devido a falta de acesso a alguns sistemas que gerem essa informação.

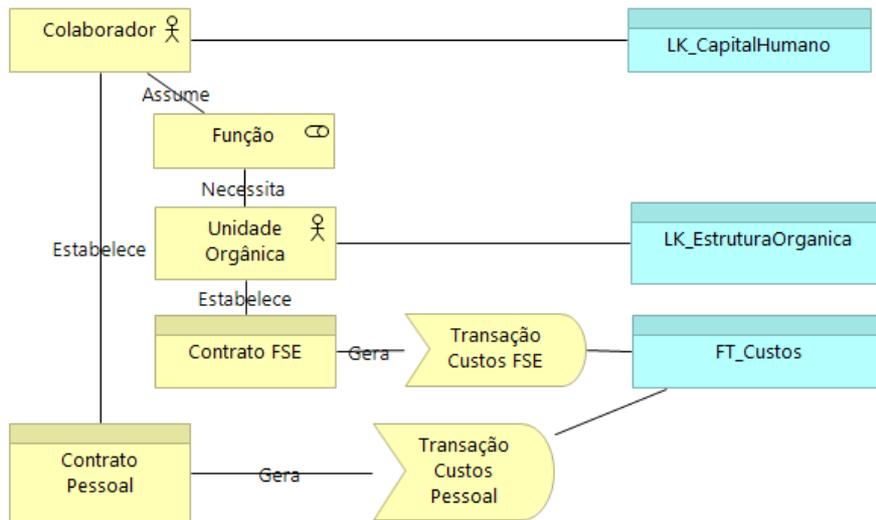


Figura 7.18: Modelo de relação entre arquitetura organizacional e arquitetura de informação de empresa gestora de infraestrutura

7.3.1.2 Arquitetura negócio

Ao nível da ontologia, sem elementos instanciados, apresentada na Figura 7.19, os canais são utilizados para comercializar oferta comercial através de contratos de proveitos estabelecidos com clientes. Os contratos têm transações associadas e que envolvem “títulos” enquanto produtos, além de “parceiros” (e.g. bolsas, *brokers*). Os clientes são tipificados por segmentos (e.g. particular, empresa).

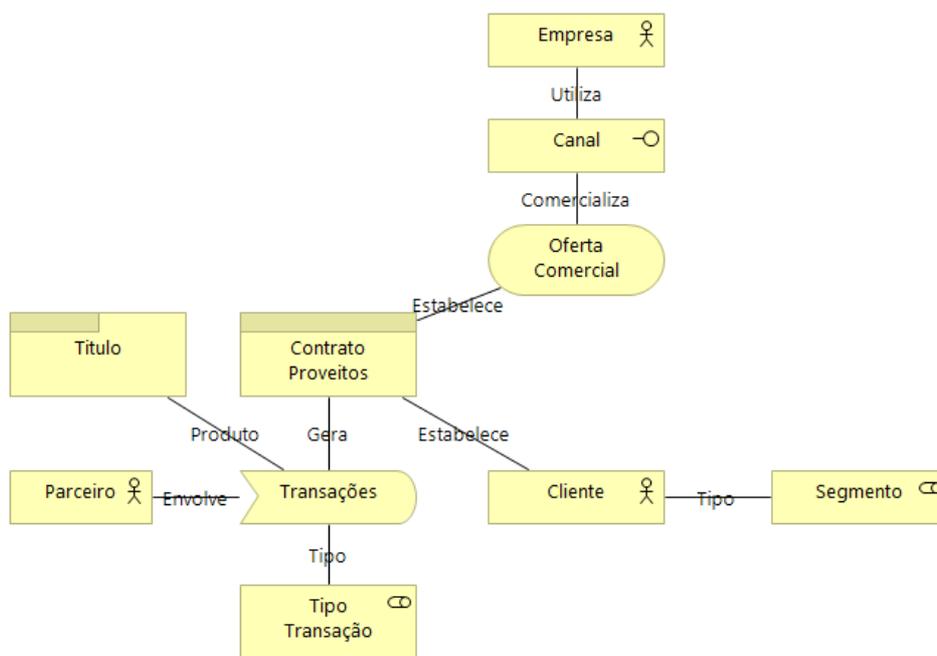


Figura 7.19: Modelo de ontologia da arquitetura negócio de empresa financeira

O modelo instanciado numa lógica de arquitetura empresarial é apresentado na Figura 7.20, para mais tardar se comparar com a visualização via *dashboard*. Não são representados elementos de lojas, parceiros, transações e clientes, pelo impacto visual complexo que tal acarretava, sendo essa uma das vantagens de se utilizar *dashboard* em *business intelligence*. O modelo de arquitetura de informação correspondente à arquitetura de negócio é apresentado na Figura 7.21., sendo de destacar o seguinte:

- Desagregação do conceito “Transações” em tabelas de dados FT_Proveitos, LK_TipoMovimento e LK_TipoEvento;
- Ausência de relações diretas entre conceitos como “Contrato Proveitos” e “Oferta Comercial” com tabelas de dados;
- Ligação do conceito “Parceiro” com tabela de dados LK_Bolsas, somente para efeito de teste, sendo que num caso mais detalhado, seria necessário representar igualmente outros parceiros como é o caso de “Brokers”.

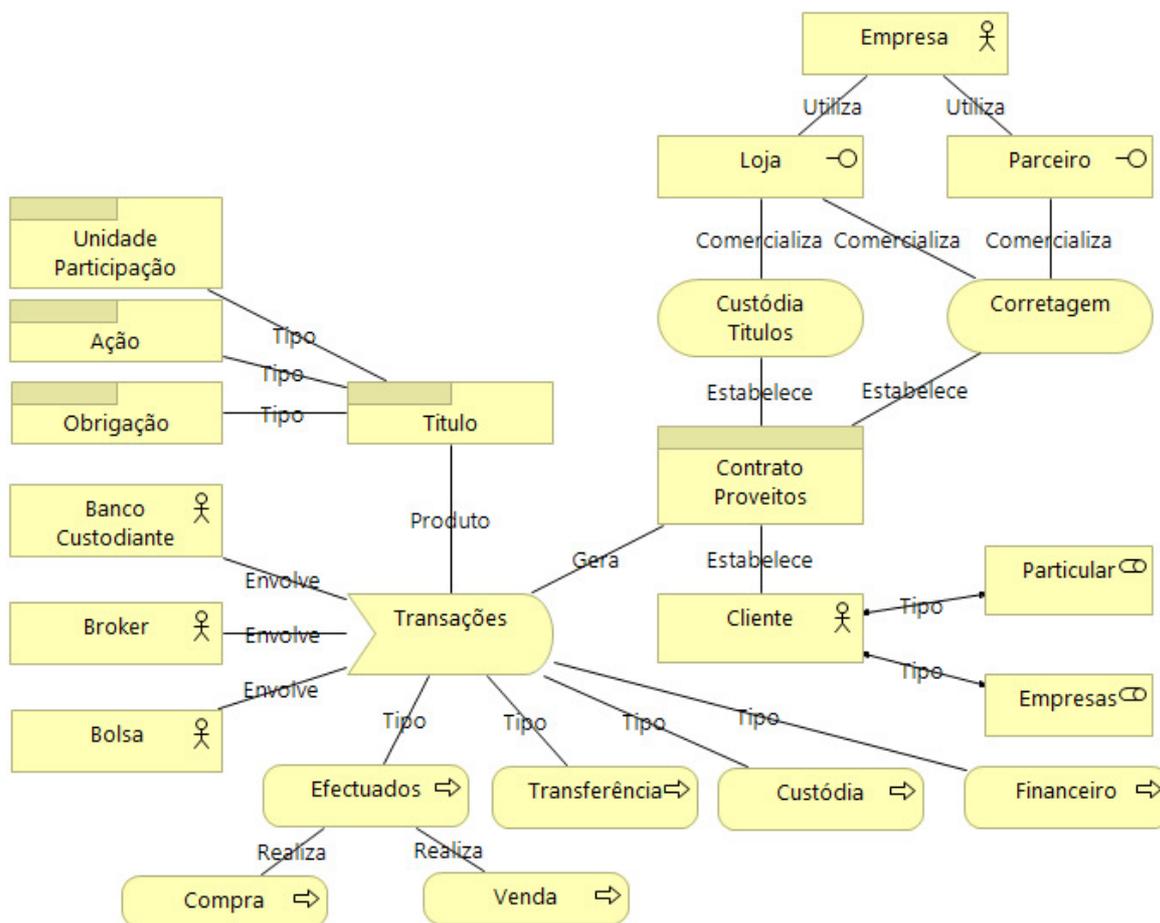


Figura 7.20: Modelo instanciado de ontologia da arquitetura negócio de empresa financeira

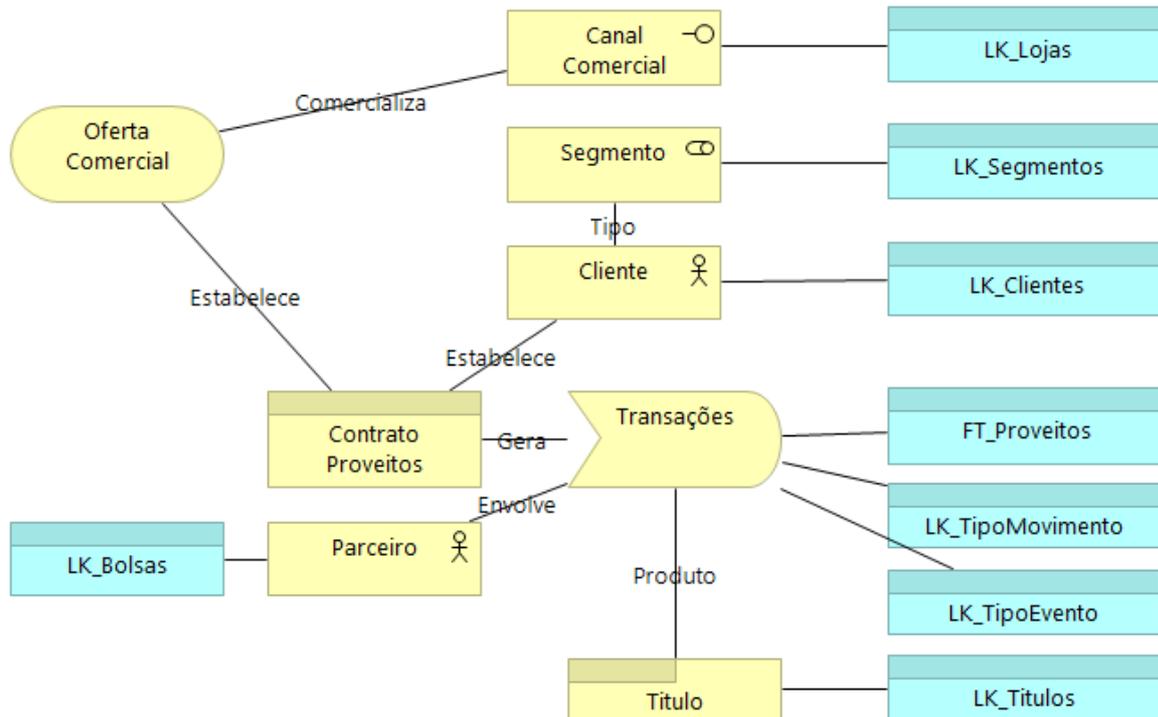


Figura 7.21: Modelo de relação entre arquitetura de negócio e arquitetura de informação de empresa gestora de infraestrutura

7.3.2 Business intelligence

7.3.2.1 Base de dados

O modelo apresentado na Figura 7.22 e relações entre tabelas apresentadas na Figura 7.23 foi criado em Microsoft PowerBI Desktop a partir de tabelas em Microsoft SQLServer. A estrutura da base de dados foi desenhada tendo por base os conceitos da organização face à definição em arquitetura empresarial sendo assim uma implementação da ontologia sobre esta forma de representação de conhecimento. Foi delimitada a factos de proveitos por não ter sido possível utilizar-se informação de custos no ambiente tecnológico utilizado para investigação. O modelo é caracterizado pelo seguinte:

- Centralidade da tabela FT_Proveitos a partir do qual se organiza um modelo em *snowflake* com dimensões como cliente (LK_Cliente) que por sua vez tem uma dimensão específica de segmento (LK_Segmento), localização geográfica pela morada (LK_Geografia) ou loja a que está associado (LK_Lojas);
- Definição de tabelas de custos, como é o caso de unidade orgânica (LK_EstruturaOrganica) e colaboradores (LK_CapitalHumano), mas sem tabela de factos associada a custos. Neste caso a ausência dessa tabela deve-se à impossibilidade de se obter os dados em causa como parte do caso de estudo com a organização.

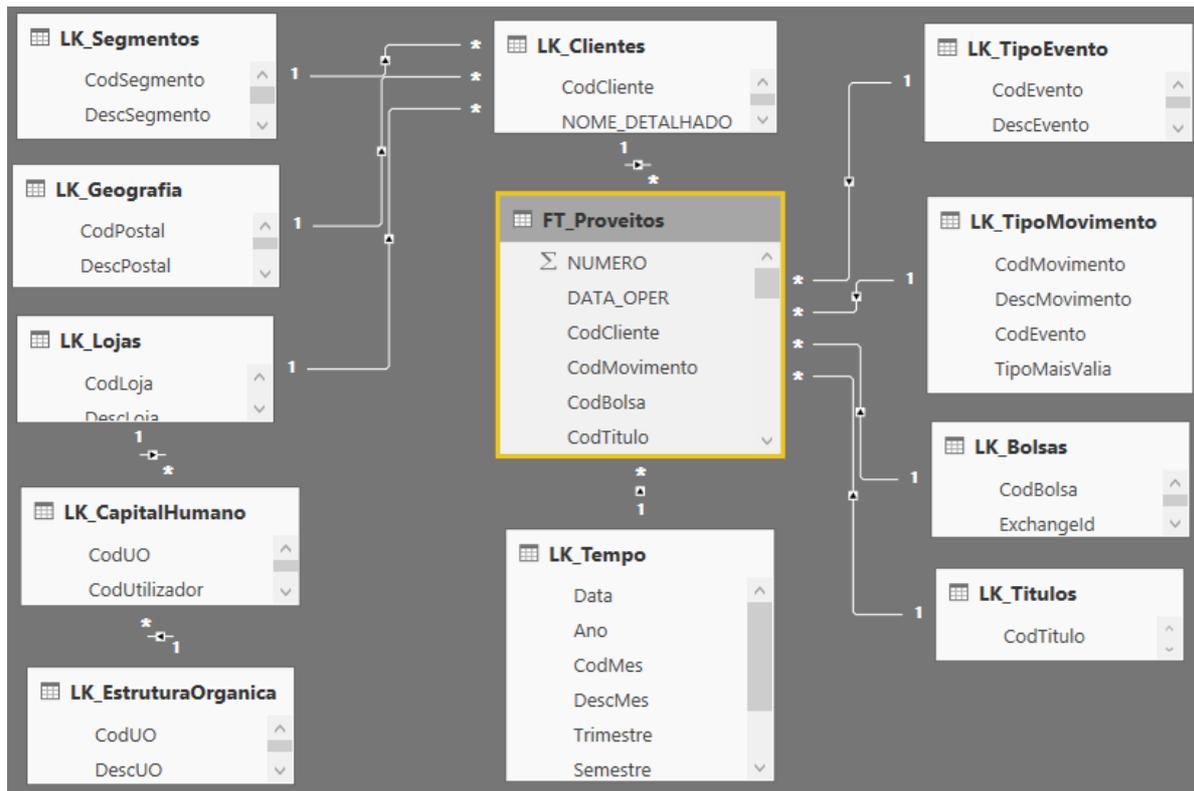


Figura 7.22: Modelo de dados empresa financeira em PowerBI Desktop

Manage relationships

Active	From: Table (Column)	To: Table (Column)
<input checked="" type="checkbox"/>	FT_Proveitos (CodBolsa)	LK_Bolsas (CodBolsa)
<input checked="" type="checkbox"/>	FT_Proveitos (CodCliente)	LK_Clientes (CodCliente)
<input checked="" type="checkbox"/>	FT_Proveitos (CodEvento)	LK_TipoEvento (CodEvento)
<input checked="" type="checkbox"/>	FT_Proveitos (CodMovimento)	LK_TipoMovimento (CodMovimento)
<input checked="" type="checkbox"/>	FT_Proveitos (CodTitulo)	LK_Titulos (CodTitulo)
<input checked="" type="checkbox"/>	FT_Proveitos (DATA_OPER)	LK_Tempo (Data)
<input checked="" type="checkbox"/>	LK_CapitalHumano (CodLoja)	LK_Lojas (CodLoja)
<input checked="" type="checkbox"/>	LK_CapitalHumano (CodUO)	LK_EstruturaOrganica (CodUO)
<input checked="" type="checkbox"/>	LK_Clientes (CodLoja)	LK_Lojas (CodLoja)
<input checked="" type="checkbox"/>	LK_Clientes (CodPostal)	LK_Geografia (CodPostal)
<input checked="" type="checkbox"/>	LK_Clientes (CodSegmento)	LK_Segmentos (CodSegmento)

New... Autodetect... Edit... Delete

Close

Figura 7.23: Modelo de relação entre tabelas da empresa financeira

Como resultado, o modelo deste caso de estudo é caracterizado pelo seguinte:

- Tabela central de factos FT_Proveitos associada a dimensões críticas de clientes (LK_Clientes), transações (LK_Eventos e LK_Movimentos), produtos (LK_Titulos) e parceiros (LK_Bolsas);
- Tabela de clientes (LK_Clientes) como dimensão, mas com uma relação com segmentação (LK_Segmento), localizações (LK_Geografia) e canal de lojas (LK_Lojas) pois de acordo com a ontologia de negócio os clientes são a base destas relações;
- Estrutura orgânica (LK_Estrutura Orgânica) e sua relação com colaboradores (LK_CapitalHumano);
- Utilização de tempo (LK_Tempo) como forma de introduzir o conceito de histórico no modelo, como é normal em modelos informacionais. Neste caso foi colocado só ao nível de tabelas de factos, apesar da modelação poder incluir também esta dimensão em tabelas de dimensões.

7.3.2.2 Integração de dados

O processo de ETL (*extraction, transform and load*) foi criado utilizando-se a ferramenta Microsoft *Integration Services*. Os processos foram organizados em agrupamentos com a mesma lógica de conceitos, da arquitetura empresarial tal como apresentado na Figura 7.24. Por essa razão, foram criados dois grupos, “Orgânica” e “Negócio” onde se considerou tarefas específicas de carregamento por tabela. As tabelas transversais como LK_Geografia e LK_Tempo foram colocadas fora destes agrupamentos.

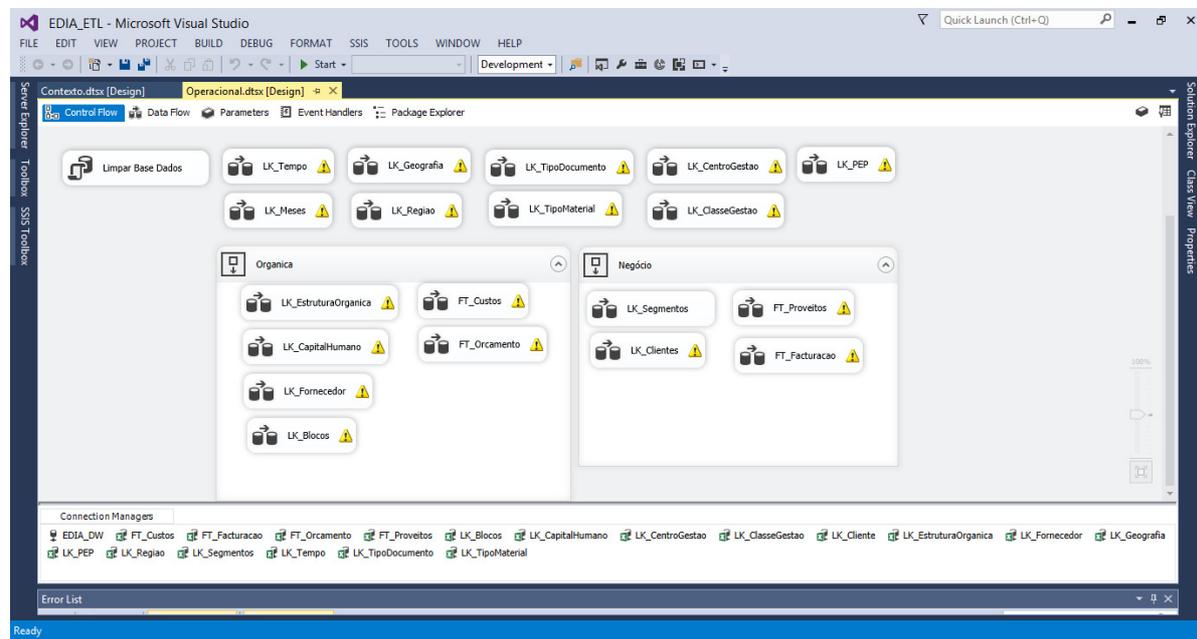


Figura 7.24: Processo ETL empresa financeira

Por outro lado, tal como apresentado na Figura 7.25, para cada tarefa existe um processo de ligação a tabelas origem (*extraction*), seguido de uma transformação (*transform*) e por fim o carregamento na tabela destino (*load*). Neste caso não foram efetuadas muitas transformações pois o modelo de base já tinha os campos preparados para carregamento.

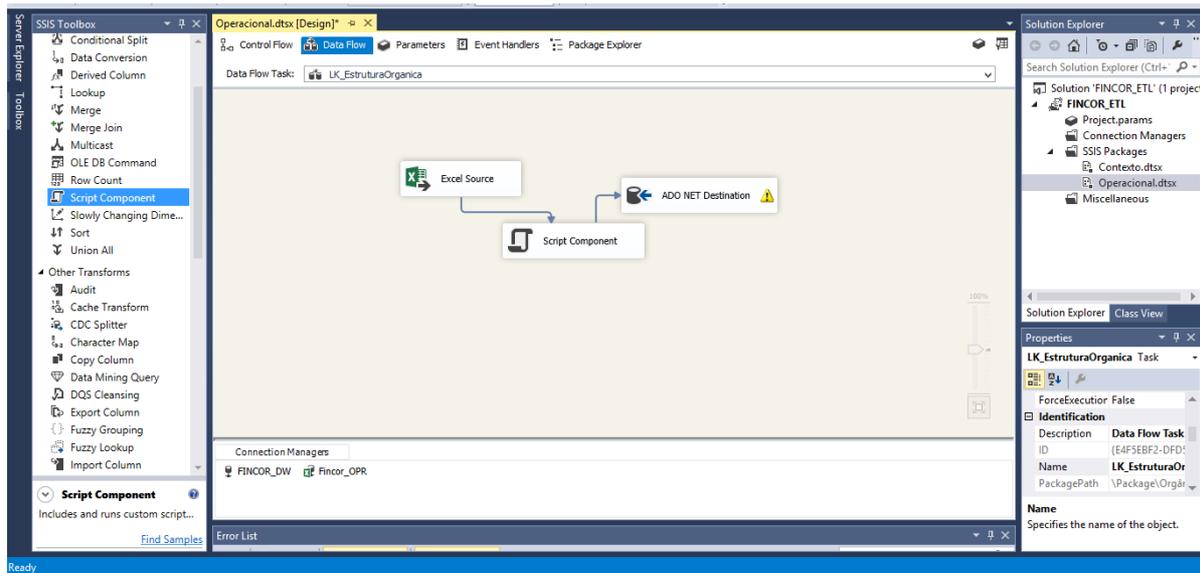


Figura 7.25: Processo ETL empresa financeira. Exemplo de transformação por tabela

7.3.2.3 Dashboard para exploração de dados

Face à orientação da arquitetura de hipótese de solução, foi criado um *dashboard* com três *sheets* relativas a “Estrutura organizacional”, “Estrutura negócio” e “Proveitos” enquanto áreas temáticas de análise da organização. Ficou em falta a *sheet* de “Custos” por falta de dados. De notar que os dados apresentados são fictícios visto terem sido mascarados a partir de dados originais da empresa em caso de estudo. Na *sheet* “Estrutura Organizacional” na Figura 7.26, pode-se observar a estrutura orgânica representada num gráfico “*heat map*” para permitir “*drill*” a partir da hierarquia orgânica sendo que o tamanho das zonas do gráfico representa a quantidade de colaboradores. Por outro lado, face aos locais de atividade da empresa, é apresentado sobre a forma de um mapa as zonas do país com o tamanho das “*bolas*” relacionada com a quantidade de colaboradores. São assim apresentados os conceitos de estrutura orgânica, colaboradores e locais, que fazem parte da modelação em arquitetura empresarial.

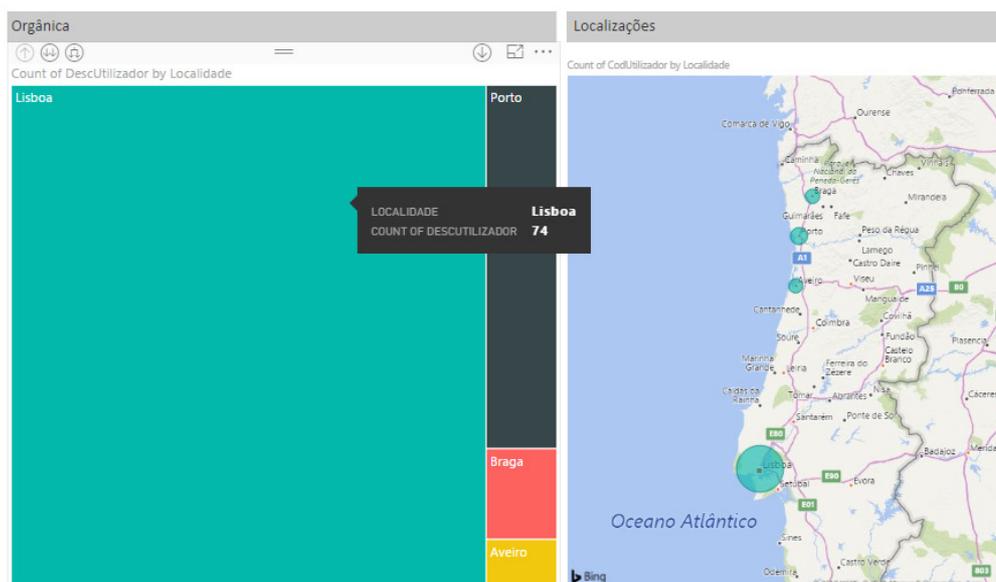


Figura 7.26: Sheet “Estrutura Organizacional” do dashboard de empresa financeira

Na sheet “Estrutura Negócio” na Figura 7.27, pode-se observar a representação de parceiros (e.g. Bolsas), produtos (e.g. serviços por tipologia de movimentos e produtos enquanto títulos comercializados), canais e segmentação de clientes.

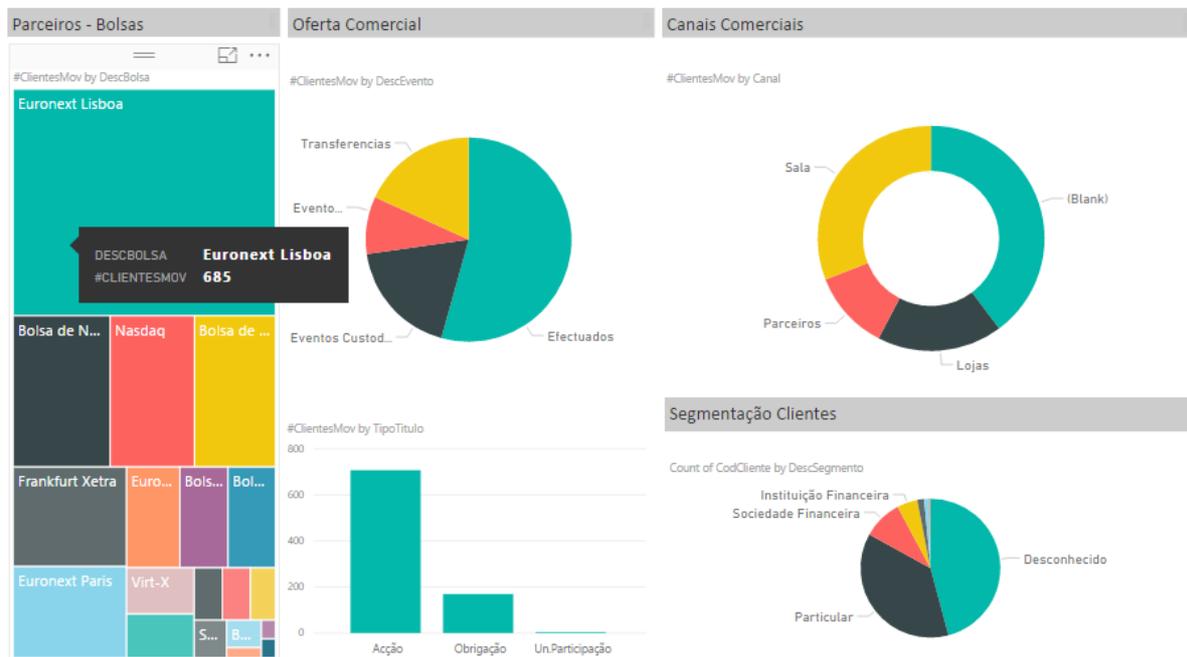


Figura 7.27: Sheet “Estrutura Negócio” do dashboard de empresa financeira

Na sheet “Proveitos” na Figura 7.28, pode-se observar que são reutilizadas as dimensões ao nível de “arquitetura de negócio”, mas acrescentado somente uma visão de tempo associado às transações e quantidade de transações.

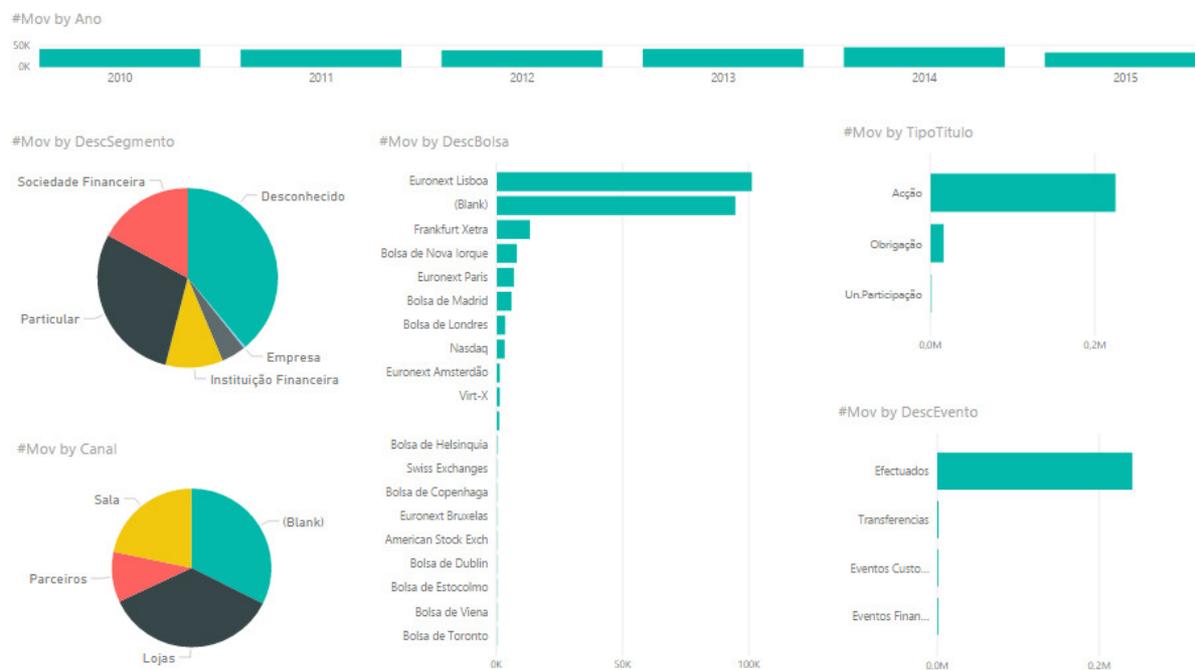


Figura 7.28: Sheet “Proveitos” do dashboard de empresa financeira

7.3.3 Analisador de expressões de negócio

Tendo como base a ontologia, é possível antes de mais detetar na visualização de *dashboard* que as expressões “métricas por dimensões” é nativa na própria ferramenta, como se repara na legenda que a ferramenta coloca em gráficos (e.g. #Mov por CodEvento; #Mov por TipoTitulo). Testando algumas expressões com a experimentação de Processamento de Língua Natural reutilizando a ontologia criada, temos um resultado adequado, tal como apresentado na Figura 7.29 para a expressão “Analisar ValorTotal, ValorUnitario de Efectuados por CodBolsa, CodCliente para CodBolsa = Euronext.”:

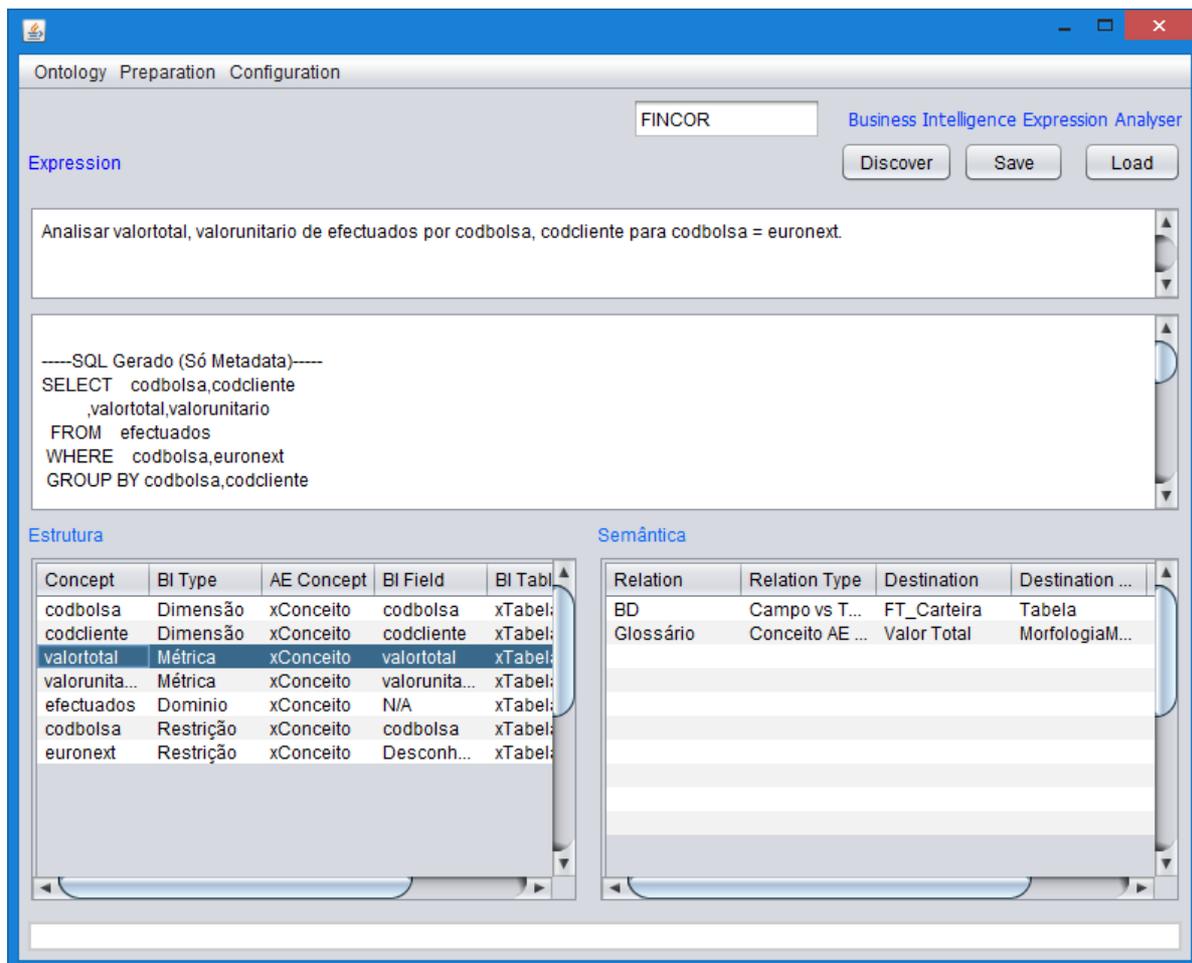


Figura 7.29: Analisador de expressões de negócio de empresa financeira

7.4 Questionário de casos de estudo

7.4.1 Utilização de arquitetura empresarial

No caso da não utilização de arquitetura empresarial, indicada na Figura 7.30, num dos casos de estudo é indicado que o facto de se documentar processos, arquitetura de sistemas de informação, organograma e plano estratégico, seria suficiente. Noutro caso é indicado que “É complexa de fazer e implica disponibilidade temporal e de procedimentos para a manter”. No entanto, é de notar que os processos, organograma, aplicações e componentes do plano estratégico (e.g. visão, missão, objetivos, produtos, canais, segmentação), fazem parte da arquitetura empresarial.

Porque razão não existe Arquitetura Empresarial (AE)

Documentação de processos é suficiente	1	50%
Documentação de SI é suficiente	1	50%
Organograma é suficiente	1	50%
Plano estratégico é suficiente	1	50%
Other	1	50%

Figura 7.30: Porque razão não existe arquitetura empresarial (AE)?

Sobre as unidades orgânicas a envolver na gestão dos conceitos relacionados com arquitetura empresarial, tal como apresentado na Figura 7.31, existe concordância sobre a necessidade de envolver as unidades orgânicas de planeamento estratégico, organização e sistemas de informação.

Que Unidades Orgânicas deviam estar envolvidas na gestão da AE ou documentação equivalente?

Planeamento Estratégico	2	100%
Organização e Métodos	2	100%
Sistemas de Informação	2	100%
Other	0	0%

Figura 7.31: Que unidades orgânicas deviam estar envolvidas na gestão da AE?

No caso dos níveis de arquitetura de conceitos relacionados com arquitetura empresarial, apresentado na Figura 7.32, destaca-se a indicação num dos casos sobre a importância do BSC (*Balance Scorecard*), visto ser um modelo utilizado para avaliação de desempenho, tal como apresentado no sub-capítulo 3.2.4. Nos restantes conceitos, destaca-se igualmente que só num dos casos de estudo é relevante a definição da orgânica e da componente tecnológica (infraestrutura de sistemas, redes e comunicações).

Que níveis de arquitetura deveriam existir no seu modelo em AE ou documentação equivalente?

Negócio sem BSC	0	0%
Negócio com BSC	1	50%
Orgânica	1	50%
Processos ou Funcional	2	100%
Aplicacional	2	100%
Informacional	2	100%
Tecnológica	1	50%
Other	0	0%

Figura 7.32: Que níveis de arquitetura deveriam existir no seu modelo em AE?

Face à não sistematização de modelação de arquitetura empresarial referida na Figura 7.20, é normal a resposta de não utilização de ferramentas na Figura 7.33, indicando-se somente num dos casos que se utilizam normativos internos. Pela análise do investigador, esses normativos estão documentados via processadores de texto e folhas de cálculo.

Que ferramenta é utilizada para documentar arquiteturas da empresa?

Normativos internos

Figura 7.33: Que ferramenta é utilizada para documentar arquiteturas da empresa?

Face à ausência de ferramentas, no caso de integração com outras fontes, tal como apresentado na Figura 7.34, é natural a ausência de outras ferramentas com as quais seria necessário integrar, destacando-se somente num dos casos a relação necessária com catálogos de produtos/serviços, por causa da captura da definição da estrutura de oferta comercial. No outro caso é indicado que “Não utilizamos atualmente uma ferramenta para a arquitetura empresarial”.

A ferramenta tem integração com outras fontes?

Sistemas de BPM/Workflow	0	0%
CMDB (Configuration Management DB)	0	0%
SOA	0	0%
Metadados BI	0	0%
Gestão de Capital Humano	0	0%
Catálogo de Produtos/Serviços	1	50%
Other	1	50%

Figura 7.34: A ferramenta tem integração com outras fontes?

Sobre a dificuldade na atualização da documentação relacionada com arquitetura empresarial, em ambos os casos de estudo é indicado como “elevado”, tal como apresentado na Figura 7.35.

Qual o nível de dificuldade na atualização da documentação?

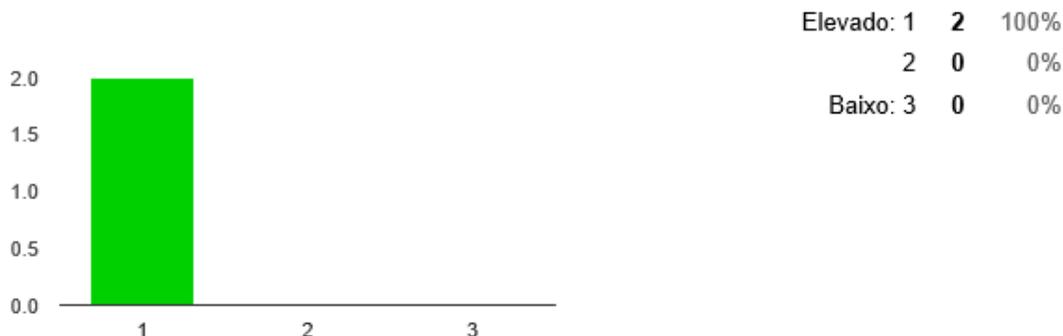


Figura 7.35: Qual o nível de dificuldade na atualização da documentação?

Considerando o nível de dificuldade elevado para atualização da documentação, os motivos apresentados na Figura 7.36, mostram uma concordância relativa ao “envolvimento da organização” e “alterações constantes de negócio”, sendo de destacar que não é indicado como motivo o alinhamento de conceitos entre negócio e sistemas de informação, bem como a ausência de dificuldade de obter informação de gestão para KPI’s. Num dos casos foi acrescentado como motivo as “alterações regulamentares”, por ser um caso de instituição financeira onde a supervisão é intensiva e com várias alterações ao longo do ano.

Qual o principal motivo na dificuldade de documentação?

Envolvimento da organização	2	100%
Obtenção de informação de ativos da organização	1	50%
Obtenção de informação de gestão para KPI's	0	0%
Alterações constantes de negócio	2	100%
Alinhamento de conceitos entre negócio e SI	0	0%
Other	1	50%

Figura 7.36: Qual o principal motivo na dificuldade de documentação?

Sobre a utilização da documentação, destaca-se essencialmente a sua relação com os processos de negócio, ou num dos casos, a sua relação com análise de impacto em alterações para sistemas de informação, tal como apresentado na Figura 7.37.

Que utilização faz da documentação de arquiteturas?

Em fase de planeamento estratégico de negócio	0	0%
Em fase de planeamento estratégico de SI	0	0%
Na análise de impacto de alterações em SI	1	50%
Na documentação de processos de negócio	2	100%
Na análise de risco de sistemas de informação	0	0%
Other	0	0%

Figura 7.37: Que utilização faz da documentação de arquiteturas?

7.4.2 Utilização de *business intelligence*

Relativamente aos níveis de arquitetura de *business intelligence*, tal como apresentado na Figura 7.38, é indicada essencialmente a *data warehouse* e modelos de exploração via *reporting* fixo e *dashboard*.

Que níveis de arquitetura de BI existem?

ODS	0	0%
DataWareHouse	2	100%
DataMart	0	0%
Reporting Fixo	1	50%
Reporting em Self-Service	0	0%
Dashboard	1	50%
BigData Analytics	0	0%
Data Mining	0	0%
Function Analytics (Profitability, Customer Intelligence, Risk Intelligence, outros)	0	0%
Other	0	0%

Figura 7.38: Que níveis de arquitetura de BI existem?

Sobre as unidades orgânicas que gerem a arquitetura de *business intelligence*, tal como apresentado na Figura 7.39, não foram selecionadas as opções em questionário. Num dos casos de estudo é indicado que é a Direção de Sistemas de Informação e noutro caso de estudo é indicado que é o *back-office* sob coordenação direta do COO (*Chief Operational Officer*).

Que Unidades Orgânicas gerem a arquitetura de BI?

Unidade específica de BI	0	0%
Unidade de gestão de aplicações de SI	0	0%
Marketing	0	0%
Planeamento Estratégico	0	0%
Other	2	100%

Figura 7.39: Que unidades orgânicas gerem a arquitetura de BI?

No caso de metamodelos utilizados, tal como indicado na Figura 7.40, num dos casos é indicado que é desenhado pela organização e noutro caso é indicado que se fez uma adaptação de modelos da ferramenta utilizada.

Que (meta)modelo é utilizado para BI?

Desenhado pela organização	1	50%
Adaptação de modelo do fabricante da ferramenta	1	50%
Other	0	0%

Figura 7.40: Que metamodelo é utilizado para BI?

Apesar de terem uma *data warehouse*, alimentada por ferramentas de ETL e com modelos de exploração de dados, em ambos os casos, tal como apresentado na Figura 7.41, é indicado que não são utilizados metadados.

Que tipo de metadata existe?

Não existe	2	100%
Gerada e utilizada no ETL	0	0%
Gerada e utilizada pelas ferramentas de exploração de BI	0	0%
Metadata transversal para ETL e Exploração BI	0	0%
Metadata Corporativa para BI e sistemas operacionais	0	0%
Other	0	0%

Figura 7.41: Que tipo de metadata existe?

Sobre as ferramentas, apresentado na Figura 7.42, num dos casos é indicado que se utilizam Microsoft (Macros e Excel), sendo que pela análise do investigador se constata que são importações de dados para Microsoft Excel a partir dos sistemas centrais, sendo depois utilizadas *Pivot Tables*. No outro caso, por análise do investigador, constata-se que se utiliza Microsoft Excel mas igualmente Microsoft *Integration Services* e Microsoft PowerBI, tendo por uma *data warehouse* em Microsoft SQLServer.

Que ferramentas são utilizadas para BI (ETL, Metadata e Exploração BI)?

Macros, Excel

Figura 7.42: Que ferramentas são utilizadas para BI?

Na análise sobre a dificuldade de gestão da arquitetura de *business intelligence*, num dos casos é indicado como elevado e noutro caso é indicado como médio, tal como apresentado na Figura 7.43.

Qual o nível de dificuldade na gestão de BI?

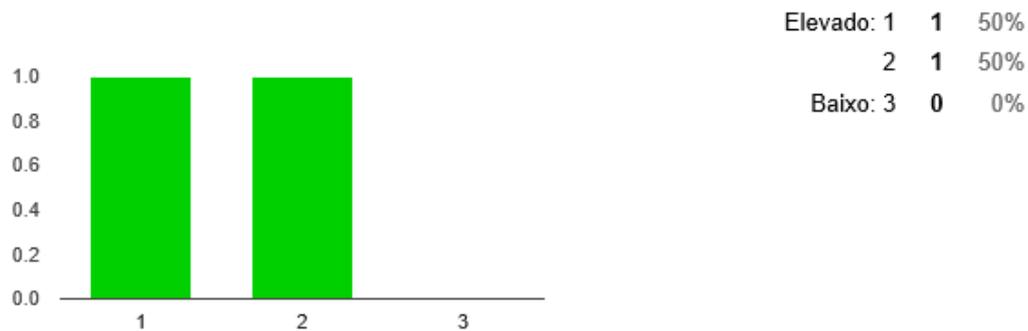


Figura 7.43: Qual o nível de dificuldade na gestão de BI?

Sobre o motivo da dificuldade de indicado no ponto anterior, tal como apresentado na Figura 7.44, os principais motivos têm a ver com a necessidade constante de incorporar novas necessidades.

Qual o principal motivo de dificuldade de gestão de BI?

Velocidade de necessidades de utilizadores para novos Dashboard	0	0%
Velocidade de necessidades de utilizadores para novos Reporte's	1	50%
Captura de novas métricas e dimensões a partir de sistemas operacionais	1	50%
Alterações de sistemas operacionais sem acompanhamento pelo BI	1	50%
Manutenção de metadados consistentes entre conceitos de negócio e conceitos de BI	1	50%
Other	0	0%

Figura 7.44: Qual o principal motivo de dificuldade de gestão de BI?

7.4.3 Experimentação da hipótese de solução

Para uma resposta às questões, foi enviado um documento resumo da tese para cada responsável das organizações envolvidas nos casos de estudo, com a seguinte estrutura em termos de capítulos e respetiva relação com os capítulos da tese aqui descrita:

- **Capítulo 1:** introdução enquanto resumo do objetivo da investigação, tal como apresentado igualmente no capítulo 1 da tese;
- **Capítulo 2:** caracterização do problema, tal como apresentado igualmente no capítulo 4 da tese, seguido da apresentação das causas e impactos do problema;
- **Capítulo 3:** descrição da hipótese de solução, tal como apresentado igualmente no capítulo 6 da tese;
- **Capítulo 4:** resultado do questionário a nove empresas tal como apresentado igualmente no capítulo 5 da tese;
- **Capítulo 5:** conclusão preliminar, tal como apresentado igualmente no capítulo 9 da tese.

Tendo por base a leitura do documento enviado que resume a experimentação que foi sendo feito ao longo de vários meses de interação com cada entidade, nos parágrafos seguintes, apresentam-se as respostas ao questionário.

No caso da opinião sobre a descrição das causas e impactos do problema, existe total concordância, tal como apresentado na Figura 7.45.

Qual a sua opinião sobre as causas e impacto do problema indicado no capítulo 2.3



Concordo	2	100%
Não concordo	0	0%
Other	0	0%

Figura 7.45: Qual a sua opinião sobre as causas e impacto do problema indicado?

No caso da opinião sobre a hipótese de utilizar a arquitetura empresarial como forma de alinhar conceitos sobre métricas e dimensões a serem consideradas em *business intelligence*, existe total concordância, tal como apresentado na Figura 7.46.

Qual a sua opinião sobre utilizar a arquitetura empresarial como base para definir métricas e dimensões em BI?

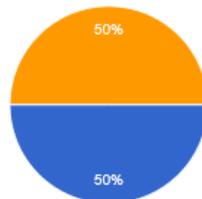


Concordo	2	100%
Não concordo	0	0%
Other	0	0%

Figura 7.46: Qual a sua opinião sobre utilização da arquitetura empresarial para definir métricas e dimensões em BI?

Sobre a comparação do resultado do questionário a nove organizações, para comparar com a realidade da organização do caso de estudo, existe total concordância, tal como apresentado na Figura 7.47. Num dos casos, é indicado, no entanto que “Concordo na globalidade”, em vez de “concordo”.

Qual a sua opinião sobre o resultado do inquérito no capítulo 4.2 face à sua organização?



Concordo	1	50%
Não concordo	0	0%
Other	1	50%

Figura 7.47: Qual a sua opinião sobre o resultado do questionário ao mercado?

Na percepção de que a modelação da arquitetura organizacional feita pelo investigador, face à análise da organização, a opinião é de total ou quase total concordância de que representa a percepção da organização, tal como apresentado na Figura 7.48.

A arquitetura organizacional apresentada reflete a realidade da sua organização?

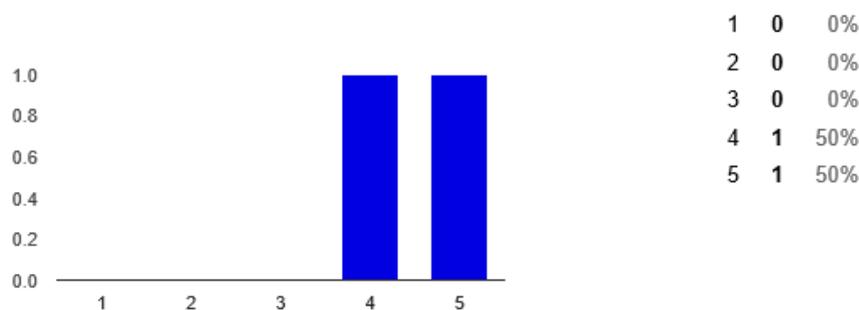


Figura 7.48: A arquitetura organizacional apresentada reflete a realidade da sua organização?

No caso da representação da arquitetura de negócio, desce ligeiramente a perceção de que representa a realidade da organização, tal como apresentado na Figura 7.49. Como nota do investigador, este facto deve-se a que não foram representados todos os conceitos, mas somente os mais críticos.

A arquitetura de negócio apresentada reflete a realidade da sua organização?

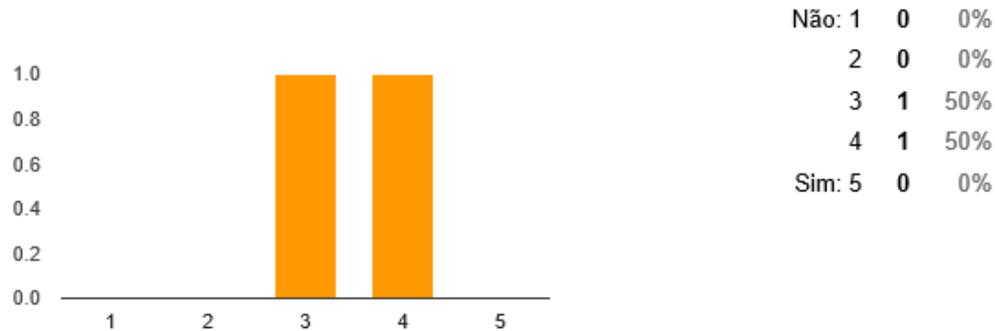


Figura 7.49: A arquitetura de negócio apresentada reflete a realidade da sua organização?

Sobre a adequação do modelo de dados de suporte à informação de gestão, criado pelo investigador especificamente para cada caso, existe uma opinião de concordância quase total de que representa as dimensões e métricas críticas da organização, tal como apresentado na Figura 7.50. Como nota do investigador, este facto deve-se a que apesar de não se ter representado a totalidade de dimensões e métricas (caso de métricas financeiras por questões de confidencialidade), as que foram utilizadas representam a estrutura central do modelo da organização.

A base de dados apresentada reflete as dimensões e factos críticos da sua organização?

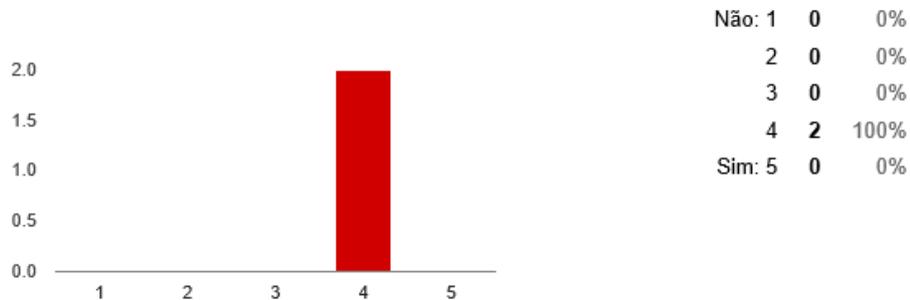


Figura 7.50: A base de dados apresentada reflete as dimensões e factos críticos da sua organização?

No caso de *dashboard* criados pelo investigador para cada caso, existe uma quase total concordância que representa um modelo de exploração adequado e alinhado com a arquitetura empresarial definida, tal como apresentado na Figura 7.51. Como nota do investigador, o facto de não ser uma concordância total, prende-se com o facto de não estarem representados todas as dimensões, métricas e perspetivas de análise.

Os Dashboard apresentados refletem a estrutura de análises (sem a totalidade das métricas) críticas da sua organização alinhada com a arquitetura empresarial?

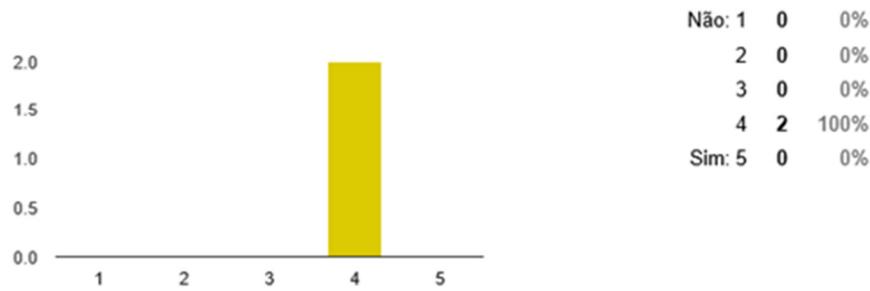


Figura 7.51: Os *dashboards* apresentados refletem a estrutura de análises críticas?

Considerando a experimentação de análise de expressões de informação de gestão e sua comparação com a exploração em *dashboard*, existe uma concordância quase total de que corresponde a uma linguagem adequada para interrogação de dimensões e métricas em *business intelligence*, tal como apresentado na Figura 7.52.

O analisador de expressões apresentado corresponde a uma linguagem de interrogação de dimensões e métricas em BI?

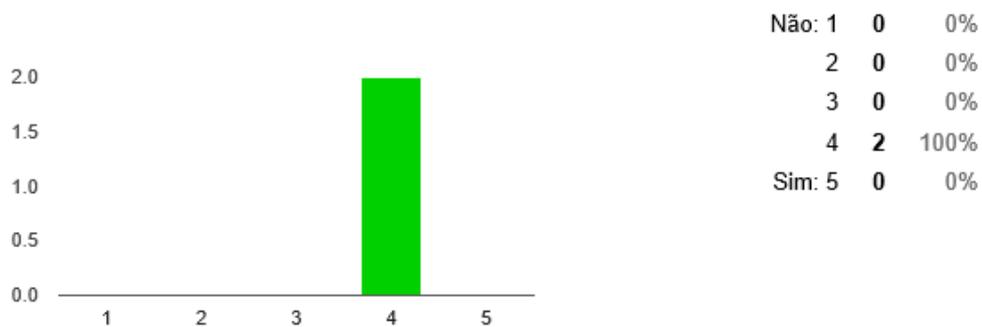


Figura 7.52: O analisador de expressões apresentado corresponde a uma linguagem de interrogação de dimensões e métricas em BI?

8 Discussão

Neste capítulo efetua-se a discussão da investigação, tendo por base a experimentação efetuada (sub-capítulo 8.2), resposta a questões de investigação (sub-capítulo 8.3) além de se analisar de forma comparativa a investigação face a outras referências (sub-capítulo 8.4).

8.1 Introdução

Considerando o problema de interoperabilidade centrado a hipótese de solução através de utilização de ontologias de domínio integrando arquitetura empresarial e sistema de *business intelligence*, efetuou-se a investigação a dois níveis. Ao nível de teste da hipótese e verificação do problema, efetuou-se experimentação prática em dois casos de estudo. Ao nível de referências equivalentes, complementou-se a análise do estado da arte com análise de investigações potencialmente equivalentes.

Na resposta às questões em investigação (sub-capítulo 8.3), teve-se em conta o seguinte:

- A arquitetura empresarial e arquitetura do sistema de *business intelligence* definidas, partiram de modelos definidos como proposta na nossa investigação para permitir de forma metodológica alinhar os conceitos na implementação. É uma experimentação que permite aferir sobre a forma como se pode modelar sistemas de *business intelligence*, partindo de um modelo de arquitetura empresarial, sem qualquer interoperabilidade técnica, mas somente recorrendo a abordagem metodológica, permitindo igualmente comparar formas de visualização da informação da arquitetura utilizando os mecanismos de exploração em arquitetura empresarial, comparando com mecanismos de exploração em *business intelligence*, neste caso com *dashboard*;
- Os modelos de arquitetura empresarial, *business intelligence*, bases de dados e glossários, geram conceitos e relações, que podem ser consolidados numa ontologia. Como tal, foi possível efetuar uma experimentação de interoperabilidade técnica, centrada num modelo de metadados como ontologia, consolidando o modelo de arquitetura empresarial, bases de dados e glossários para classificação de métricas e dimensões, além de descoberta da semântica da informação. É uma experimentação delimitada, no sentido em que o objetivo não foi a geração automática de métricas e dimensões num modelo de *business intelligence*, mas somente a experimentação de formas de integrar os conceitos num sistema de metadados para efeito de exploração dos mesmos no contexto de descoberta de conhecimento;
- Considerando a ontologia de domínio criada, a experimentação permitiu igualmente aferir sobre a adequação de uma linguagem de informação de gestão baseada em análise de expressões utilizando processamento de língua natural.

No caso da análise comparativa com outras investigações de referência, teve-se em conta que não existem casos equivalentes, pelo que se optou por comparação com as investigações em componentes parcelares, face à nossa investigação. Em particular, como referido por Mochol (2009), a interoperabilidade é um problema considerado em várias investigações no domínio da web semântica no sentido em que se torna necessário trocar informação entre os vários sistemas mantendo um entendimento da semântica da informação.

Na identificação de casos de referência equivalentes à nossa investigação considerou-se as seguintes pesquisas sobre linhas de investigação focadas em ontologias:

- **Ontologias em modelação:** como se tem utilizado o conceito de ontologias, utilizando OWL aplicado a arquitetura empresarial, UML e outras formas de capturar o modelo da organização e dos sistemas de informação?
- **Ontologias em bases de dados:** considerando que as ontologias utilizam linguagens de representação de dados semânticos, qual a utilização destas linguagens em sistemas de gestão de bases de dados que possa permitir criar modelos de integração direta entre modelos de bases de dados já em OWL com modelos de representação do modelo de negócio da organização igualmente em OWL?
- **Ontologias em *business intelligence*:** pelo facto das dimensões e métricas serem a base de modelos de informação de gestão ao nível da representação dos modelos de dados informacionais e na exploração da informação, qual a utilização das ontologias ao nível de definição de metadados neste domínio?
- **Ontologias em metadados:** considerando a classificação de informação através de taxonomias, “folksonomias” e metadados, qual a utilização de ontologias neste domínio que permita inferir sobre a sua utilização para anotações e classificação de dados tendo por base ontologias organizacionais definidas como arquitetura empresarial ou de outras formas, incluindo em *data governance*?

O modelo investigado baseado na linguagem de notação ArchiMate foi focado na captura das relações entre conceitos em AE e dimensões e métricas em BI, tendo-se para tal apresentado o mapeamento na secção 6.3 como parte da arquitetura de solução. Como tal, não se considerou a semântica de relações de vários tipos por não ser o foco de investigação, motivo pelo qual se optou por associações gerais entre conceitos em AE face ao objetivo de capturar somente a relação entre conceitos e não o tipo de relação entre conceitos. Por outro lado, face a limitações de elementos para identificar métricas em modelação em AE, considerou-se a utilização de elementos de AE como eventos, associado ao conceito organizacional de originador de transações, enquanto proposta devidamente justificada na arquitetura de solução.

Para a modelação de AE investigou-se a relação entre o modelo de Osterwalder e Pigneur (2010), utilizado hoje em organizações para posicionamento estratégico, e níveis arquiteturais de AE, tendo-se definido um modelo da investigação na secção 4.2 como ponto de partida. Não consideramos nesta aproximação que todos os modelos devem estar na AE (e.g. modelo de dados, modelo de processos, layout de edifícios, modelos de posicionamento de produtos). A base de entendimento foi de que a AE deve ter um catálogo de primeiro nível de todos os conceitos enquanto visão holística que permite capturar conceitos e relações, motivo pelo qual é a base de uma ontologia organizacional, argumentando-se desta forma que se são uma ontologia, então deve estar suportado em linguagens como OWL para permitir maior interoperabilidade entre sistemas, o que não seria possível se fosse utilizado somente as estruturas próprias de ferramentas como Archi ou outras ferramentas de modelação de AE. No caso específico de modelação de processos utilizand-se linguagens como BPMN (Business Process Modelling Notation), estas não foram consideradas por estarem fora do âmbito do que era o foco da investigação nesta fase.

8.2 Experimentação da hipótese

8.2.1 Modelação de sistemas de *business intelligence* a partir da arquitetura empresarial

Dos casos de estudo considerados na investigação deteta-se um padrão face ao problema identificado associada à interoperabilidade:

- Não existe uma estrutura de metadados reutilizada entre sistemas. Cada sistema tem os seus próprios metadados sem uma interface passível de integração;
- O alinhamento de conceitos entre o modelo da organização e sistemas de informação é efetuado por esforço de análise de requisitos.

Ao nível da hipótese de solução concluímos o seguinte:

- A abordagem de modelação de arquitetura empresarial considerando a arquitetura organizacional e de negócio para enquadrar a relação entre conceitos como uma ontologia, criando um mapeamento por tipo de componentes utilizados para relacionar com dimensões e métricas, resultou como adequada. É possível notar nos casos utilizados a reutilização de conceitos e relações ao nível de arquitetura empresarial, e a sua visualização em *business intelligence* ao nível de modelação, integração e exploração de dados;
- Torna-se mais fácil identificar a partir deste alinhamento qualquer alteração necessária ao nível de conceitos na arquitetura empresarial e seu impacto ao nível de dimensões e métricas identificadas em *business intelligence*, e vice-versa;
- O modelo de exploração de dados em *business intelligence* a partir de *dashboard*, utilizando gráficos, tabelas e mecanismos de navegação com “*Drill*”, permite uma melhor representação de conceitos instanciados com muitos elementos. Como tal consegue-se representar estruturas base de conceitos (e.g. Segmento-tem-clientes é representado como quantidade e detalhe de clientes por segmentos) e mesmo hierarquias mais detalhadas (e.g. unidades orgânicas é representado por quantidade e detalhe de colaboradores por direção e departamento). Permite igualmente utilizar componentes gráficas para filtrar outros componentes (gráficos e tabelas), o que é um acrescento de funcionalidade para navegação pelos dados. Por último, evita a necessidade de integração de KPI's (isto é, métricas) na arquitetura empresarial pois em *business intelligence* estes já estão utilizados como parte da própria visualização.

8.2.2 Utilização de ontologias em *business Intelligence* para descoberta de conhecimento

O protótipo aplicado aos casos de estudo, permitiu integrar sem qualquer limitação, as seguintes fontes de informação:

- **Bases de dados:** tendo por base a sua definição em DDL conseguiu-se traduzir os conceitos de tabelas, campos, formatos de campos e tipos de chave face a uma correspondência com a ontologia ao nível de *class*, *objet properties* e *data properties*. Desta forma foi possível representar em OWL o modelo de dados;

- **Glossários:** considerando que os campos têm nomes técnicos e por sua vez cada campo como conceito tem várias definições, foi possível testar o carregamento de conceitos equivalentes e relacionar com os campos, integrando na ontologia comum;
- **Conceitos arquitetura empresarial:** considerando os conceitos, propriedades e relações definidas no modelo em ArchiMate, foi possível gerar para OWL e integrar na ontologia comum, estendendo e reconciliando desta forma os conceitos organizacionais.

Com base na ontologia criada foi possível experimentar com sucesso o seguinte:

- Utilizar inferência para classificar o que são métricas e dimensões;
- Utilizar inferência para enriquecer a semântica do modelo, permitindo com base em cada atributo, enquanto classe da ontologia, entender as suas propriedades e classes equivalentes. Esta possibilidade permite explorar de forma adequada o conceito de descoberta de dados e conhecimento, no sentido em que os conceitos e relações correspondem a conhecimento sobre os metadados organizacionais, que de outra forma obrigava a reconciliar de forma mais onerosa e muitas vezes manual, os vários conceitos. Permite igualmente olhar para esta informação numa lógica de auditoria de informação focada em *data governance* corporativa.

8.2.3 Utilização de ontologias em *business intelligence* via processamento de língua natural

O protótipo aplicado aos casos de estudo, tendo por base a ontologia de domínio, permitiu aferir a adequação da solução de analisar expressões de informação de gestão tendo por base a ontologia e processamento de língua natural.

Com base na solução implementada em laboratório, as expressões analisadas têm a seguinte forma: “ANALISAR <Lista de métricas> POR <Lista de domínios> POR <Lista de dimensões> PARA <Lista de restrições>”. É como tal uma expressão fechada onde se posiciona cada conceito, que de acordo com a sintaxe da frase, permite aplicar técnicas de processamento de língua natural de *sentence detector*, *tokenizer* e *part-of-speech*, para delimitar cada palavra.

Sendo uma expressão fechada, a técnica *part-of-speech* foi utilizada para identificar algumas palavras no segmento de frase (e.g. sujeitos, artigos, números), mas a identificação de métricas ou dimensões recorreu a pesquisa por SPARQL e algoritmos em Java utilizando a ontologia como fonte de dados. Como tal, na identificação de cada palavra não se seguiu técnicas de processamento de língua natural ao nível de *name entity recognition* baseado em dados de treino como lista de palavras classificadas, pelo facto de se ter optado por utilizar o modelo de ontologia resultante da inferência.

O resultado obtido permitiu verificar a adequação da solução para várias expressões, mostrando ao utilizador as dimensões, métricas, domínios e restrições identificadas, permitindo para cada conceito detalhar a semântica de cada termo nos metadados, isto é, com base em pesquisa sobre a ontologia para enquadramento do mapeamento entre o termo (isto é, palavra) e sua localização como conceito (AE, Glossário e DDL) e suas relações.

8.3 Resposta a questões de investigação

Face à questão central de investigação podemos considerar o seguinte:

Como utilizar metadados como ontologias para garantir a interoperabilidade entre um sistema de arquitetura empresarial e um sistema de business intelligence? Utilizando o modelo criado em ArchiMate, enriquecido com glossários e estruturas de bases de dados utilizados em business intelligence, para se criar um modelo de metadados corporativo visto como ontologia.

Face às questões complementares podemos considerar o seguinte:

- ***Como implementar um sistema de business intelligence a partir do modelo definido numa arquitetura empresarial?*** Reutilizando os mesmos conceitos definidos em arquitetura empresarial por áreas específicas (e.g. organizacional, negócio), criando tabelas, processos de integração e formas de exploração, utilizando os mesmos conceitos face a equivalências entre os mesmos, como parte de uma abordagem metodológica;
- ***Qual a relação entre conceitos definidos em arquitetura empresarial e dimensões e métricas definidos em business intelligence?*** As dimensões correspondem diretamente a conceitos base e a relações entre conceitos na arquitetura empresarial. As métricas correspondem a contagem por conceitos e a propriedades em conceitos, nomeadamente eventos como transações, contratos e produtos, onde residem propriedades que correspondem a valores financeiros;
- ***Como se reutilizar a ontologia criada para responder a questões de utilizadores via processamento de linguagem natural, no contexto de knowledge discovery?*** Considerando que a ontologia integra o glossário torna-se possível utiliza-la para entender cada palavra numa expressão e daí interpretar o que são atributos, métricas e domínios da informação, permitindo uma forma de descoberta de conhecimento sobre metadados, além de permitir a geração de *queries* DML automaticamente a partir de questões em linguagem natural.

Das conclusões, foi possível ainda estruturar respostas sobre as seguintes questões metodológicas:

- ***Qual a relação entre modelo da organização e sistemas de informação?*** O modelo da organização corresponde a um conjunto de conceitos e relações entre si, que têm uma correspondência direta com modelo de dados (tabelas, campos), modelo de dimensões e métricas, que são a base dos sistemas de informação de gestão em particular;
- ***Qual o papel das dimensões e métricas na relação entre modelo da organização e modelo de sistemas de informação ao nível de business intelligence?*** As dimensões e suas hierarquias são derivadas dos componentes utilizados para modelar as

organizações enquanto as que métricas resultam de eventos e processos enquanto conceitos de transações que derivam factos mensuráveis para o negócio;

- **Qual a causa da heterogeneidade e falta de interoperabilidade em torno de um modelo comum da organização entre modelação da organização e sistemas de informação?** A causa está relacionada com a impossibilidade de garantir que sistemas vistos como *packages* com tecnologia e fabricantes diferentes e implementados em fases diferentes do ciclo da vida das organizações sejam passíveis de integrar sem existirem normas. Por outro lado, as várias tentativas de criar estas normas não têm resultado pela falta de adesão dos fabricantes e desenvolvedores de soluções tecnológicas;
- **Qual o benefício de reutilizar uma solução de metadados transversal implementada como ontologia em business intelligence?** Permitiria utilizar uma espécie de glossário comum para a toda a organização, facilitando a criação de uma linguagem comum e maior eficácia e eficiência na implementação dos sistemas.

Face à experimentação em laboratório com ArchiMate, OWL, ApacheJENA, OpenNLP e Microsoft BI, foi possível igualmente responder a um conjunto de questões técnicas:

- **Como classificar as métricas e dimensões a partir de arquitetura empresarial vista como ontologia explorada em business intelligence?** Recorrendo aos tipos de componentes utilizados em ArchiMate. Por exemplo, as estruturas ativas são tipicamente dimensões, enquanto componentes como eventos e processos derivam em métricas. Por outro lado, a análise gramatical das palavras permite inferir sobre conceitos que são dimensões e conceitos que são métricas, além de se poder utilizar a nomenclatura e formato de campos;
- **Como modelar a organização através de uma arquitetura empresarial para estar alinhada com o modelo de business intelligence para efeito de implementação de dados e exploração da informação com os mesmos conceitos?** Considerando o nível de arquitetura “Organização” e “Negócio” como a base para estruturar os conceitos e relações chave, colocando as métricas de custos ao nível da “Organização” e as métricas de proveitos ao nível de “Negócio”;
- **Que características deve ter um repositório para acomodar a integração de vários conceitos da organização e suportar a interoperabilidade?** Deve ser flexível para representar metadados de conceitos e relações, mas extensível ao se instanciar com elementos, permitindo criar novas relações sem alterar a estrutura, o que é algo nativo do modelo de ontologia;
- **Que tipo de tecnologia utilizar para explorar este repositório comum?** Tecnologia de modelo de dados semânticos como é o caso de OWL, em vez de RDF, pela necessidade de criar definições base e tipos de relações ao nível de *object properties* e com potencial de se expandir com regras em SWRL.

8.4 Trabalhos relacionados

8.4.1 Ontologias em modelação

Para modelação das organizações existem abordagens de utilização de ontologias ao nível de modelação de sistemas via UML (*Unified Modelling Language*) ou ao nível de linguagens de modelação de arquitetura empresarial, nomeadamente o ArchiMate (GROUP, 2012, 2016). No entanto, estas abordagens estão orientadas para a conversão de modelos para RDF/OWL e não para a integração com outros modelos ao nível de bases de dados ou sistemas informacionais. Esta realidade torna-se evidente nos seguintes casos:

- Maisie (2013) contextualiza a sua investigação num domínio de problema onde as organizações utilizam vários sistemas de informação interdependentes que tornam complexa a atividade de adaptação dos sistemas face a mudanças na organização. A sua investigação considera igualmente a arquitetura empresarial como ferramenta para modelação da organização que pode ser vista como uma ontologia, mas que apresentam o problema de não estarem relacionadas com repositórios de dados concretos nos sistemas de informação. Como tal, apresenta uma solução baseada em ArchiMEO enquanto ferramenta suportada em ArchiMate para modelação de arquitetura empresarial como ontologia, considerando que esta abordagem representa uma forma de metadados organizacionais;
- Antunes et al. (2013) e Morais (2013) referem que a arquitetura empresarial tem por objetivo alinhar o modelo de negócio com os sistemas de suporte utilizando níveis arquiteturais para capturar as várias visões interligadas da organização. No entanto, consideram que a arquitetura empresarial utiliza linguagens genéricas de meta-modelos que torna complexa a integração quando se modela cada nível de arquitetura para depois se integrar numa visão global da organização. Para o efeito, propõe a utilização de ontologias para modelar cada domínio da organização para facilitar a sua integração com consistência e rastreabilidade;
- Belghiat e Bourahla (2012) propõem uma abordagem baseada em transformação por grafos para a automatização de geração de ontologias em OWL a partir de diagramas de classes em UML;
- Thonssen (2013) investiga a utilização da arquitetura empresarial para geração de metadados complementado com uma ontologia empresarial que descreve partes do repositório de conhecimento da empresa por grupo de processos ou áreas funcionais;
- Wickett (2012) desenvolveu um referencial para categorizar coleções e detalhe de objetos para serem utilizados no desenvolvimento, controlo de qualidade e conversão de expressões técnicas de metadados em representação lógica mais próxima da linguagem do utilizador;
- Bermudez (2004) criou um modelo abstrato de representação de metadados no domínio geoespacial para permitir interoperabilidade e mapeamento semântica entre fontes distintas. No mesmo domínio, Arpinar et al. (2004) explora igualmente as ontologias aplicadas a metadados em domínio geoespacial para raciocínio, pesquisa e visualização;
- Singh (2013) desenvolveu um modelo de metadados e modelo de classificação para ser aplicado em componentes de *software* utilizando ontologias e taxonomias;

- Turco (2010) apresenta no seu livro sobre “*enterprise architecture & metadata modelling*” o conceito de *enterprise metadata repository* tendo por base uma análise das funções de negócio para centrar a temática na modelação de uma arquitetura de informação transversal para toda a organização. Não foca, no entanto, o tema da relação direta entre arquitetura empresarial, ontologias e *business intelligence*;
- Szwed (2015) desenvolveu um *plug-in* para a ferramenta de modelação de arquitetura empresarial em Archi que permite transformar um modelo ArchiMate para OWL considerando os componentes da arquitetura como *class* com as devidas hierarquias com relações pai-filho e colocando as outras relações entre componentes de diversos tipos enquanto *object properties*. Este *plug-in* foi utilizado na nossa investigação para gerar o modelo OWL a partir da modelação de arquitetura empresarial em ArchiMate;
- Rajabi et al. (2013) e Kang et al. (2010) desenvolveram modelos de ontologia empresarial alinhada com os conceitos do referencial de Zachman como hipótese de solução para o problema de inconsistência e interoperabilidade entre modelos de arquitetura empresarial e dados presentes nos sistemas de informação.

No caso específico de utilização de *business intelligence* para implementação de memória organizacional existem investigações diversas com destaque para as seguintes:

- **Aquisição de informação:** Blue et al. (2011) consideram o papel do ETL para a captura dinâmica da informação. Wei e Qing-pu (2007) posicionam o *data mining* para melhorar a eficiência da captura de informação;
- **Repositório de informação:** Sciarrone (2009) utiliza modelos multidimensionais para suporte a OLAP. Nemati et al. (2002) utilizam uma arquitetura de *knowledge warehouse* para armazenar a memória organizacional;
- **Exploração de informação:** Borrajo et al. (2011) utilizam modelos de aprendizagem automática como suporte a análise preditiva sobre casos de falhas de decisão. Farquard et al. (2009) utilizam *support vector machine* e *naive bayes tree* para extrair regras como parte de soluções de *data mining*. Chung et al. (2005) definiram um modelo visual para *knowledge discovery* para melhorar a interação com utilizadores.

8.4.2 Ontologias em bases de dados

Existem atualmente bases de dados semânticas ou *ontology database* suportadas em RDF/OWL como é o caso da GraphDB com múltiplas utilizações. Ao nível de investigação, Lependu (2010) propõe uma solução para implementação de bases de dados semânticas sobre bases de dados relacionais para reutilizar as vantagens de armazenamento, escalabilidade e exploração de grandes volumes de dados.

As referências analisadas assentam em soluções de extração da definição de bases de dados relacionais e criação de alternativas de mapeamentos para definir classes e propriedades para gerar a ontologia. Existem igualmente alternativas ao nível de linguagens de interrogação de bases de dados relacionais mas utilizando linguagens baseadas em ontologias efetuando-se mapeamentos lógicos entre os modelos. No entanto, nenhuma das soluções utiliza a informação de base da arquitetura empresarial para captura inicial de conceitos chave do negócio para otimizar o mapeamento entre as bases de dados relacionais e a ontologia, enriquecendo assim o modelo final, tal como a seguir apresentado:

- Bumans (2010) refere várias ferramentas e estudos sobre a transformação de bases de dados relacionais em bases de dados semânticas representadas em esquemas RDF ou ontologias OWL. Tendo por base as soluções existentes, propõe um mecanismo de transformação baseado em mapeamento ao nível de um esquema designado por “*meta-level relational schema*” utilizando uma linguagem de especificação de mapeamento para permitir ler bases de dados relacionais via SQL e gerar RDF/OWL ao nível de tripletos (*isto é, sujeito, predicado e objeto*);
- Villanueva-Rosales (2011) desenvolveu o *framework* DBOwlizer para desenho automático de ontologias a partir de bases de dados relacionais na terceira forma normal. O *framework* permite extrair informação de estrutura da base de dados para criar mapeamento e geração de ontologias utilizando heurísticas baseadas em regras;
- Zhou et al. (2010) desenvolveram uma solução baseada em ApacheJENA para implementar a transformação de bases de dados relacionais em ontologias OWL a partir de uma extração inicial da definição da base de dados para permitir um mapeamento utilizado na geração da ontologia;
- Chokri (2007) propõe uma abordagem de interrogação baseada em ontologias para acesso a bases de dados relacionais. Para o efeito, utiliza ontologias para transformar a solicitação lógica de um utilizador numa solicitação física face à base de dados, através de mapeamentos de conceitos tendo por base a ontologia;
- Alalwan (2011) considera o problema de distribuição de dados e heterogeneidade de definições dos mesmos no contexto de integração de bases de dados. Para resolver o problema, utiliza uma abordagem com inclusão de ontologias híbridas obtendo a semântica da ontologia local a partir dos modelos lógicos de bases de dados através de *reverse engineering*;
- Rodriguez-Muro et al. (2008) desenvolveram um *plug-in* para o Protégé para OBDA (Ontology-Base Data Access) enquanto forma de acesso a dados em bases de dados relacionais utilizando uma mediação por ontologias;
- Lubyte (2010) refere várias alternativas para descrever ontologias que efetuam mediação de acesso a fontes de dados, estabelecendo mecanismos para definir mapeamentos que relacionam as ontologias com as fontes de dados e para suporte a exploração das ontologias assim construídas. Considera no entanto que os mapeamentos estão pré-definidos, pelo que propõe um procedimento automático para gerar ontologias baseado em restrições de integridade de dados definidas nas fontes de dados relacionais utilizando *reverse engineering*.

8.4.3 Ontologias em *business intelligence*

Os sistemas informacionais podem tirar partido das ontologias pela sua necessidade de conhecerem os sistemas aplicativos origem, onde obtêm a informação. Esta necessidade deriva do facto de o início do seu processamento considerar regras de transformação para classificação de atributos e métricas de negócio e pelo facto dos metadados técnicos e de negócio estarem relacionada com dados origem e modelos de negócio da organização.

No entanto, as referências encontradas sobre utilização de ontologias e modelação organizacional reutilizada em sistemas informacionais não consideram normalmente a

existência de ferramentas de arquitetura empresarial como base do modelo representativo da organização enquanto metadados a serem reutilizados. Por outro lado, não consideram igualmente a forma como a representação de metadados como ontologias ao nível da modelação das organizações e dos metadados específicos de sistemas informacionais, podem ser conjugados em torno de uma hipótese de solução. Esta realidade torna-se evidente nos seguintes casos de investigação:

- Cao et al. (2004) utilizam ontologias em três camadas (visão do utilizador ao nível de um portal de conhecimento sobre relatórios de gestão, visão global analítica ao nível da definição de metadados do sistema de *business intelligence* e visão física correspondente às estruturas de dados dos sistemas operacionais origem que são carregados para o sistema de *business intelligence*). Utilizam *ontology mapping & query parsing* para implementar um modelo designado por “*ontology service-based match and translation*”. Não utilizam no entanto o enriquecimento de metadados a partir da arquitetura empresarial dependendo assim da estrutura e semântica da informação já residente nas bases de dados de sistemas operacionais;
- Stefanov (2007) refere que as investigações em *business intelligence* são focadas em bases de dados e não na interação entre a informação de gestão e os modelos de negócio das organizações. Por essa razão a sua investigação endereça o problema da relação entre *business intelligence* e a estrutura, comportamento e objetivos organizacionais. Para o efeito desenvolveu uma linguagem conceptual de modelação para capturar a organização e uma abordagem de metadados de negócio orientada a este modelo para relacionar objetivos e processos de negócio com modelos de dados de *business intelligence*. Não considera no entanto a existência da arquitetura empresarial para modelação e ontologias como base da solução, referindo somente que são alternativas;
- Pinto (2009) desenvolveu ontologias em Protégé sobre OWL para suporte ao processo de descoberta de conhecimento em base de dados e para o processo de *database marketing*. Kremen (2012) desenvolveu um *framework* para desenvolvimento de sistemas de informação utilizando ontologias suportadas em OWL e SPARQL. Szwed (2015) desenvolveu uma ontologia para um modelo de referência de arquitetura para *business intelligence*. Ghawi (2010) desenvolveu uma solução para o problema de cooperação entre vários sistemas utilizando ontologias criadas a partir de cada sistema utilizando mapeamento entre as fontes para suporte à tradução de interrogações entre os vários sistemas. Chapman (2005) utilizou ontologias não só para a aquisição e armazenamento, mas também para exploração da informação, combinada com ferramentas de *business intelligence*. Cheng et al. (2009) utilizaram uma abordagem de *business intelligence application* baseada em ontologias, utilizando análise estatística e *data mining* para melhorar a retenção e exploração de conhecimento organizacional. No entanto, nenhuma destas ontologias parte de um modelo da empresa ao nível de arquitetura empresarial para alinhar os seus conceitos com os que são utilizados em *business intelligence*;
- Chowdhury et al. (2013) utilizaram metadados semânticos como suporte para uma solução para responder a questões de utilizadores em sistemas de *business intelligence* gerando as necessárias interrogações sobre os sistemas. Ghosh et al. (2014) e Kaur e Bali (2012) abordaram modelos para transformar texto livre em SQL para resposta a questões de utilizadores utilizando processamento de língua natural. Saias et al. (2012)

desenvolveram uma hipótese em linguagem natural para facilitar a interação entre utilizadores e cubos OLAP, utilizando conversão para MDX. No entanto, não consideram uma ontologia de domínio no contexto empresarial baseada em modelos de negócio e reutilizada para alinhar com conceitos de dados;

- Wizdee (2016), Kueri (2016) e Microsoft English Query (2016), são casos de empresas que desenvolveram soluções para traduzir expressões em língua natural para *queries SQL*. Não existem detalhes de implementação, visto serem soluções comerciais, mas não partem de um modelo de arquitetura empresarial para definir a ontologia, reconciliável com glossários ou bases de dados.

8.4.4 Ontologias em metadados

Os metadados permitem a identificação dos dados existentes em sistemas, considerando a sua estrutura e a sua semântica, podendo agrupar-se em metadados técnicos e de negócio. Devem ser possíveis de integrar numa visão organizacional a partir de vários repositórios existentes. No entanto, existem formas de representação, níveis de detalhe e formas de utilização distintas de acordo com as ferramentas que as utilizam. Como tal, existem normas de metadados em vários domínios como o DCMI ou IMM, que apesar de pouca adesão pelos implementadores de ferramentas, permitem fazer uma aproximação de uma forma de solução para o problema. Por outro lado, existem várias investigações que partem das ontologias como elemento integrador.

No entanto, as investigações analisadas neste domínio não abordam o caso específico das organizações e em particular, não consideram a criticidade da relação entre modelo de negócio e informação centrada em conceitos chave de dimensões e métricas. Esta realidade torna-se evidente nos seguintes casos de investigação:

- Veterlli et al. (2000) investigam normas para interoperabilidade de metadados para concluir que a divergência entre *standards* são o maior obstáculo para a implementação de metadados organizacionais;
- Mochol (2009) desenvolveu um *framework* designado por MOMA (*Metadados-Based Ontology Matching*) como suporte à integração entre várias ontologias;
- Maedche e Staab (2001) desenvolveram uma abordagem para comparação de metadados onde os objetos são descritos via ontologias para permitir identificar as semelhanças na comparação. As comparações são efetuadas tendo por base as semelhanças de taxonomia, relações e atributos;
- Chen et al. (2008) desenvolveram uma solução utilizando Protégé e ApacheJENA para transformar metadados representados em XML para uma representação em RDF/OWL que permite melhorar a captura semântica dos dados, a relação com outros conjuntos de dados e a melhoria de interoperabilidade;
- Handschuh (2005) investigou a forma como se pode gerar metadados através de anotações no contexto da web semântica para criar um *framework* de anotações para reduzir a sua complexidade de utilização pelos utilizadores;
- Scilia (2006) ao analisar a semântica ou significado do recurso que é a informação conclui que a infraestrutura para representação de conhecimento tem aproveitado o conceito

de ontologias para permitir a partilha de semântica ao nível de metadados. Como tal, considera que esta abordagem multidisciplinar onde se combina as técnicas de descrição de metadados com a engenharia de ontologias introduz uma tendência de investigação multidisciplinar neste domínio;

- Mendiáldua (2011) abordou o tema das taxonomias para classificação de metadados, comparando com “folksonomias” enquanto anotações dos utilizadores no contexto de classificação de conteúdos em redes sociais. No entanto, é uma aproximação passível de ser aplicada em organizações, com ontologia de domínio assente num modelo de conhecimento de termos e conceitos da organização.

9 Conclusão

Neste capítulo são apresentadas as conclusões finais da tese. Para o efeito, resume-se o problema e arquitetura de solução (sub-capítulo 9.2), são apresentados os principais resultados alcançados e limitações (sub-capítulo 9.3), as publicações efetuadas (sub-capítulo 9.4) e as linhas de investigação futuras (sub-capítulo 9.4).

9.1 Introdução

A investigação está focada num problema de interoperabilidade entre arquitetura empresarial e *business intelligence*, investigando-se as ontologias como solução que permita a reutilização de conceitos entre os dois sistemas, unificando os metadados da arquitetura empresarial, de bases de dados e de glossários para se constituir num modelo comum de metadados, visto como uma ontologia corporativa.

Para se contextualizar o problema e elaborar-se uma hipótese de solução foi efetuado um levantamento do estado da arte (capítulo 3). Como tal, primeiro analisou-se o domínio da gestão das organizações, focando a forma como esta é estruturada para efeito de planeamento estratégico e operações, incluindo a forma como se efetua a monitorização de desempenho organizacional, com o objetivo de entender os vários modelos de gestão utilizados. De seguida analisaram-se os domínios da gestão de conhecimento e dos sistemas de informação, para se entender a relação entre o modelo da organização e a forma como este modelo é utilizado na representação de conhecimento, processamento da informação e exploração de conhecimento. Nesta análise, separou-se a análise de dados relacionados com eventos que ocorrem na organização e a análise dos dados que representam a estrutura e semântica da própria organização. Com este levantamento, foi possível identificar a forma como os sistemas de informação, em particular os sistemas de *business intelligence*, podem utilizar o modelo da organização como ontologia, para estruturarem informação para avaliação do desempenho organizacional, e qual o papel dos metadados e das ontologias como parte da solução.

Face ao estado da arte, caracterizou-se o problema (capítulo 4) e elaborou-se um questionário para nove organizações (capítulo 5) para permitir aferir sobre a relevância do problema e adequação da abordagem seguida para hipótese de solução.

Com o problema sistematizado sob a forma de requisitos para um sistema, elaborou-se uma arquitetura como hipótese de solução (capítulo 6) criando-se um conjunto de modelos e um sistema em laboratório para experimentação em dois casos de estudo (capítulo 7). Os resultados da experimentação e análise da solução face a outras referências foram apresentados como parte da discussão da investigação (capítulo 8).

9.2 Problema e hipótese solução

O problema de interoperabilidade surge entre vários sistemas numa organização face à ausência de uma estrutura de dados comum e reutilizada enquanto metadados corporativos. No entanto, focamos a investigação na interoperabilidade entre arquitetura empresarial (AE) e *business intelligence* (BI) pela necessidade de encontrar uma solução para alinhamento entre a definição do modelo da organização e a monitorização do seu desempenho.

Considerando a realidade dos dois tipos de sistemas (AE e BI), constatamos o seguinte:

- A representação do modelo da organização é efetuada como uma arquitetura empresarial, mas a monitorização de desempenho organizacional é efetuado via *business intelligence*;
- Os conceitos utilizados na arquitetura empresarial (e.g. clientes, canais, fornecedores, produtos) são equivalentes a conceitos de bases de dados utilizados para criação de tabelas e relações entre tabelas em qualquer tipo de sistema, mas têm igualmente uma correspondência com a definição de dimensões de análise e eventos críticos para identificação de métricas em sistemas de *business intelligence*;
- Ao nível da arquitetura empresarial, pode-se modelar a arquitetura de negócio que é vista como metadados de negócio para os restantes sistemas, mas deve-se igualmente modelar a arquitetura de informação enquanto metadados técnicos, com relações entre os dois níveis de arquitetura;
- Na modelação de *business intelligence*, começa-se por identificar dimensões e métricas, que correspondem a conceitos de negócio. Como tal, a linguagem de negócio de informação de gestão em *business intelligence* utiliza conceitos mais próximos da arquitetura empresarial ao nível de negócio do que de modelos de dados, pelo que se existisse uma ontologia de domínio que pudesse ser reutilizada, poderia acelerar o esforço de implementação e garantir assim um maior alinhamento entre os sistemas;
- No entanto, os sistemas de AE e BI têm formas de representação de informação distintas, com tecnologias distintas, fases de implementação diferentes ao longo da vida das organizações e não utilizam modelação da organização como metadados organizacionais reutilizados entre os vários sistemas, motivo pelo qual subsiste um problema a ser investigado.

Como não existe um modelo atual que permita a interoperabilidade entre AE e BI, em torno de metadados vistos como ontologia organizacional, cada sistema tem a sua própria estrutura de metadados e os conceitos implementados têm problemas de alinhamento de estrutura e semântica. Por outro lado, alguns sistemas são *packages* de vários fabricantes, com diferentes arquiteturas internas e implementados em várias fases das organizações, pelo que não permitem que se garanta uma integração em torno de uma estrutura de metadados transversal. As ontologias surgem assim como forma possível de integração, face à sua flexibilidade para definição de conceitos e relações, além de que permitem utilizar mecanismos de *ontology learning* e inferência com base em lógica do domínio da ontologia. O problema tem sido abordado em outras investigações de referência (sub-capítulo 8.4) mas sem uma relação entre arquitetura empresarial e *business intelligence*, sendo que os trabalhos relacionados consideram essencialmente as seguintes abordagens:

- Transformar bases de dados relacionais em bases de dados semânticas, mas considerando somente a exploração de informação nesta forma de representação de bases de dados para acesso a bases de dados relacionais via linguagens semânticas como SPARQL ou inclusive aplicando processamento de língua natural para transformar questões em texto para *queries* DML/SQL. Não permite como tal alinhar o modelo de arquitetura empresarial com *business intelligence* nem permite encontrar uma solução de correspondência para definir dimensões e métricas;

- Converter modelos de arquitetura empresarial em modelos semânticos OWL. Pode ser visto como uma forma intermédia para reutilizar estes modelos em *business intelligence* ou para ser enriquecido com os próprios metadados de *business intelligence*. No entanto, não existem referências neste sentido. Não permite como tal alinhar conceitos de arquitetura empresarial, e métricas e dimensões em *business intelligence*;
- Inferir sobre classificação de métricas e dimensões a partir de classificação de campos em bases de dados tal como é efetuado em sistemas de *business intelligence*. Não existem investigações sobre este tema, sendo que no entanto, as ferramentas comerciais, e em *open source*, utilizam o formato dos campos para o efeito.

Neste sentido, consideramos o seguinte enquadramento para hipótese de solução:

- Ambos os sistemas utilizam metadados para representação de dados sobre a estrutura e semântica de conceitos da organização. No caso da AE, a definição da organização pode ser ela mesma vista como metadados de negócio ou como ontologia corporativa;
- Ambos os sistemas utilizam os mesmos conceitos, mas com perspetivas diferentes. Na AE o objetivo é estruturar um modelo da organização considerando a sua estratégia. No sistema de *BI* o objetivo é acompanhar o negócio e suas tendências utilizando informação de acordo com métricas (medição de factos do negócio) e dimensões (estruturas base de organização), mas numa perspetiva tácita e estratégica;
- É necessário complementar a ontologia criada na AE com informação de bases de dados e glossários utilizados em *BI*. Na base do alinhamento está o entendimento e classificação de métricas e de dimensões, utilizadas em metadados técnicos em sistemas (regras de negócio, tabelas, campos) e em metadados de negócio (conceitos de negócio, estratégia e glossários);

Tendo por base o problema, considerou-se o seguinte princípio de solução:

- A AE deve somente capturar a estrutura e relação de conceitos, como metadados ou ontologia, em vez de instanciar estes conceitos. Deve-se considerar ao nível da AE um modelo específico de arquitetura de informação, relacionado com a arquitetura de negócio e organizacional, mas visto de forma independente da arquitetura aplicacional, para permitir identificar os principais conceitos de dados enquanto informação crítica da organização. Este nível de arquitetura permite estabelecer a relação entre conceitos de negócio e tabelas no modelo de dados;
- Os conceitos definidos na AE podem ser relacionados com dimensões e métricas em *BI* de acordo com o tipo de cada conceito em AE ou por interpretação do significado das palavras, utilizando processamento de língua natural;
- Deve ser criada uma ontologia por integração de conceitos definidos em AE, com conceitos definidos em bases de dados (via DDL) e glossários. O sistema de *BI* deve reutilizar esta ontologia para implementar as suas bases de dados e processos de ETL, mantendo os mesmos conceitos tal como definidos na ontologia;
- Com este alinhamento, obtém-se um primeiro resultado prático relacionado com a exploração da informação de gestão com uma navegação explícita da componente “Organização” e “Negócio” em visualizações de *dashboard*, reutilizando as mesmas dimensões em análises específicas de “Custos” e “Proveitos”, que antes eram vistas em

sistemas de navegação em AE. Este resultado deriva do facto de que em AE a navegação de detalhe (e.g. segmentos de um cliente, tipologia de produtos) é menos adequada do que a sua exploração em modelos de BI utilizando funcionalidades como *drill* (detalhar por dimensões). Por esta razão, na investigação começamos por posicionar a AE para definir somente os conceitos base e suas relações, mas sem instanciar os conceitos;

- A mesma ontologia, vista como conceitos de métricas e dimensões, pode ser reutilizada em sistemas de processamento de língua natural para permitir uma interação com utilizadores com base numa ontologia ou linguagem comum na organização.

Considerando o enfoque da tese na investigação da interoperabilidade entre arquitetura empresarial e *business intelligence*, foi definida a seguinte questão principal:

Como utilizar metadados como ontologias para garantir a interoperabilidade entre arquitetura empresarial e business intelligence?

Esta questão central foi complementada com as seguintes questões:

Como implementar um sistema de business intelligence a partir do modelo definido numa arquitetura empresarial?

Qual a relação entre conceitos definidos numa arquitetura empresarial e dimensões e métricas em business intelligence?

Como se pode reutilizar a ontologia criada para responder a questões de utilizadores via processamento de linguagem natural, no contexto de knowledge discovery?

9.3 Resultados alcançados e limitações

Considerando o problema definido, foi formulada uma arquitetura enquanto hipótese de solução a partir de uma escolha de cenários que foi objeto de validação pela seguinte via:

- Revisão da literatura, que permitiu verificar a adequação da hipótese de solução mediante resposta a questões teóricas e técnicas inicialmente colocadas;
- Realização de um questionário a nove empresas de vários sectores de atividade para entender o contexto de utilização de AE e de BI, como ponto de partida para validação da adequação da hipótese formulada e da importância do problema;
- Realização de testes em laboratório tendo por base casos de estudo com duas organizações, onde foi formulada de raiz um modelo de AE delimitado aos níveis de arquitetura relevantes, para servir de base a uma implementação de um sistema de BI delimitado a um processo de integração de dados, modelo de dados e exploração por *dashboard*. Foi ainda testado o analisador de expressões de processamento de língua natural e de metadados. Os resultados permitiram validar a estabilidade do modelo de solução e comparar a visualização do modelo da organização via arquitetura empresarial ou utilizando *dashboard* de *business intelligence*. No final foi realizado um questionário com cada organização dos casos de estudo, para aferir sobre as conclusões da experimentação.

Face à experimentação, análise de questionários e revisão da literatura, podemos concluir o seguinte:

- Não existe uma investigação totalmente equivalente, nem existem sistemas em *open source* ou comerciais que utilizem a mesma abordagem utilizada na nossa investigação;
- O modelo base da organização deve ser centralizado em definições de estruturas de “Organização” e “Negócio” onde se enquadram as dimensões base e as métricas necessárias para avaliar o desempenho organizacional em termos de “Custos” e “Proveitos” financeiros, respetivamente, que permitem montar sistemas de análise de rentabilidade. Esta arquitetura de negócio deve estar relacionada com uma arquitetura de informação onde constam os principais conceitos de dados, sendo que esta arquitetura deve ser independente da arquitetura aplicacional;
- O modelo de “Organização” como parte da arquitetura de negócio, permite detalhar as dimensões relacionadas com fornecimentos e serviços externos, orgânica e recursos humanos, que é a base do modelo de “Custos” utilizado em sistemas de *BI*;
- O modelo de “Negócio” como parte da arquitetura de negócio, permite detalhar as dimensões relacionadas com transações de venda de produtos e serviços, que é a base do modelo de “Proveitos” utilizado em sistemas de *BI*;
- As dimensões utilizadas em informação de gestão correspondem a conceitos utilizados em *AE*, pelo que a sua identificação é mais facilitada. No entanto, as métricas correspondem a dados derivados de conceitos específicos de eventos (e.g. transações comerciais) definidos na *AE*, pelo que é uma identificação mais indireta;
- Ao se modelar na *AE* a visão detalhada da organização e a visão como metadados vistos como ontologia, além de se implementar os conceitos como dimensões e métricas em *BI*, torna-se possível uma comparação da exploração do modelo da organização como metadados numa arquitetura empresarial e em alternativa pelo modelo de navegação com gráficos, tabelas, “*Drill*” e relação entre componentes em *dashboard de business intelligence*. Desta análise conclui-se que o modelo com *dashboard* permite apresentar mais detalhe da instância de conceitos, tornando-se assim a forma mais adequada de apresentar o modelo da organização;
- A análise de expressões do tipo “ANALISAR <lista de métricas> DE <lista de domínios> POR <lista de dimensões> PARA <lista de restrições>” permitiu concluir sobre a adequação do modelo construído com processamento de língua natural enquanto solução para interação com utilizadores, e na localização de informação constante em metadados associada a interrogações de negócio. Poder-se-ia, no entanto, retirar da frase a expressão “DE <lista de domínios>” pois a interpretação de métricas e dimensões, permite automaticamente identificar os domínios em causa.

As principais limitações na investigação foram as seguintes:

- Ausência de modelos equivalentes para efetuar comparações de resultados diretos. Este facto deve-se a que a maioria de investigações equivalentes está focada na conversão de modelos de dados relacionais para OWL, na conversão da AE para OWL ou na definição de taxonomias;
- Impossibilidade de testar a reutilização do repositório de metadados como ontologia como parte integrante de uma solução de *BI*, gerando-se automaticamente a estrutura de métricas e dimensões. Para se experimentar esta extensão de solução, seria necessário utilizar um sistema de *BI* com metadados abertos, o que só seria possível com sistemas em *open source* e após se fechar as conclusões desta investigação. Como tal, foi considerado como investigação futura;
- Impossibilidade de integrar na ontologia os metadados dos processos de ETL e visualizações em *dashboard*, pois para tal, ter-se-ia que estar dependente de modelos específicos de cada ferramenta de ETL e solução de visualização de dados em *BI*, exceto se essas ferramentas exportassem para SQL, e nesse caso, poderia utilizar-se *reverse engineering* para obter as *queries* DML por trás das transformações (ETL) ou na visualização (MDX ou outra linguagem OLAP). Como seria necessário fechar primeiro o modelo desta tese, foi considerado como investigação futura.

9.4 Publicações efetuadas

Durante o período de investigação entre novembro de 2013 e setembro de 2016 foram publicados vários artigos em Portugal e Espanha, com o objetivo de aferir sobre a adequação dos temas em investigação, tal como apresentados na Tabela 9.1:

Título	Publicação	Ano
Enterprise Intelligence based on Metadata and Enterprise Architecture Model Integration.	Jornadas Informática da Universidade de Évora, 2013.	2013.
Inteligência empresarial suportada em metadados vista como ontologia organizacional.	Jornadas Informática da Universidade de Évora, 2014.	2014.
Implementação de metadados como ontologia.	Conferência CISTI2015, 2015.	2015.
Auditoria de informação baseada em ontologia.	Conferência WICTA2015, 2015.	2015.
Enterprise Intelligence based on ontology metadata.	Conferência Doctoral Consortium JCSD2015, 2015.	2015.

Tabela 9.1: Publicações efetuadas durante a investigação

9.5 Linhas de investigação futuras

A investigação apresenta os seguintes contributos:

- Posicionamento de metadados como ontologias para interoperabilidade entre *BI* e *AE*, mantendo a arquitetura empresarial o papel de identificar, classificar e relacionar conceitos chave, como base de uma ontologia, definindo um modelo *base de data governance* corporativa com uma representação em linguagens de ontologias como *OWL* para facilitar a interoperabilidade entre sistemas;
- Exequibilidade e limitações de utilização de ontologias como forma de implementar metadados, considerando integração por *ontology learning* e utilização de inferência como forma de *data lineage*;
- Acelerador para metodologias de implementação de sistemas de *BI* face à identificação e classificação de dimensões e métricas a partir da *AE*;
- Utilização de processamento de língua natural sobre ontologias para analisar expressões da organização que permitam criar novas formas de interação com sistemas de *BI*, *AE* e *data governance*.

Estes contributos abrem as seguintes perspetivas de investigação futuras alinhadas com os resultados desta investigação:

- Estender o modelo de metadados desenvolvido, para integrar processos de *ETL* e visualizações em *dashboard* a partir de interpretação de *DML*. Permite fechar o modelo unificado de metadados como ontologia, integrando normas como *ArchiMate*, *DDL* e *DML* para criar uma ontologia comum de metadados técnicos e de negócio e possibilitando *data lineage* via inferência no contexto mais abrangente de *Data Governance*;
- Criação de mecanismos para deteção automática de novas dimensões e métricas mediante novas definições em bases de dados ou como resultado de modelação em *AE*;
- Reutilização de modelo de ontologia como metadados para unificar ferramentas de modelação de *AE* e de *BI*, partilhando assim um único modelo de *data governance* e passando a ser um sistema de *enterprise intelligence*, permitindo a sua reutilização por outros tipos de sistemas para além de *BI*;
- Criação de mecanismos para comparar organizações por sector de atividade ou por áreas funcionais, tendo por base a estrutura das ontologias de domínio sectorial;
- Permitir a integração com sistemas de visualização de dados para contextualizar métricas e dimensões de acordo com a semântica de dados utilizados na visualização. Possibilita igualmente a reutilização por qualquer outra aplicação de sistemas de informação para contextualizar os dados geridos nessas aplicações. Representa uma forma de glossário no contexto da utilização da informação;
- Utilização de agentes inteligentes sobre a ontologia para personalizar a relação dos utilizadores com a informação de gestão de que necessitam, quando necessitam e com que suporte para a tomada de decisão, com reutilização em processos de negócio ou aprendizagem automática.

Referências bibliográficas

- Abecker, A., Bernardi, A., Ntioudis, S., van Elst, L., Herterich, R., Houy, C., e Müller, S. (2001). Workflow-embedded organizational memory access: the DECOR project. *IJCAI'2001 Workshop on Knowledge Management and Organizational Memories*, 63-74. doi:10.1007/978-1-4615-0947-9_6.
- Adomavicius, G., e Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734-749. doi:10.1109/tkde.2005.99.
- Alalwan, N. (2011). *Ontological Approach for Database Integration*. Doctoral Thesis, Faculty of Technology De Montfort University, United Kingdom, England.
- Angeloni, M. T., e Fernandes, C. B. (2000). Organizações de conhecimento: dos modelos à aplicação prática. *Encontro de estudos organizacionais ENEO*, Curitiba: GEO/ANPAD.
- Angles, R., e Gutierrez, C. (2008). Survey of graph database models. *ACM Computing Surveys*, 40(1), 1-39. doi:10.1145/1322432.1322433
- Antunes, G., Caetano, A., Bakhshandeh, M., Mayer, R., & Borbinha, J. (2013). Using Ontologies to Integrate Multiple Enterprise Architecture Domains. *Business Information Systems Workshops Lecture Notes in Business Information Processing*, 61-72. doi:10.1007/978-3-642-41687-3_8
- ApacheOpenNLP. (2015). Disponível em <https://opennlp.apache.org/>, último acesso a 07/02/2015.
- Arnett, D. B., Menon, A., e Wilcox, J. B. (2000). Using competitive intelligence: Antecedents and consequences. *Competitive Intelligence Review*, 11(3), 16-27. doi:10.1002/1520-6386(200033)11:3<16::aid-cir4>3.0.co;2-z.
- Arpinar, I., Sheth, A., Ramakrishnan, C., Usery, L., Azami, M., e Kwan, M. (2004). Geospatial Ontology Development and Semantic Analytics. *University of Georgia*.
- Astrova, I. (2009). Rules for mapping SQL relational databases to OWL ontologies. *Metadata and Semantics (pp. 415-424)*. Springer US.
- Barbieri, C. (2001). *BI – Business Intelligence: Modelagem e Tecnologia*. Rio de Janeiro: Axcel Books.
- Belghiat, A., e Bourahla, M. (2012). UML Class Diagrams to OWL Ontologies: A Graph Transformation based Approach. *International Journal of Computer Applications*, 41(3), 41-46. doi:10.5120/5525-7566.
- Bellandi, A., Nasoni, S., Tommasi, A., e Zavattari, C. (2010). Ontology-Driven Relation Extraction by Pattern Discovery. *2010 Second International Conference on Information, Process, and Knowledge Management*. doi:10.1109/eknow.2010.17.
- Berkeley. (2006). UC-Berkeley Data Warehouse Roadmap: Data Warehouse Architecture. Disponível em <http://docplayer.net/2545057-Uc-berkeley-data-warehouse-roadmap-data-warehouse-architecture.html>, último acesso a 01/05/2016.
- Bermudez, L. (2004). *ONTOMET: Ontology Metadata Framework*. Doctoral Thesis, Drexel University.

- Berners-Lee, T., Weitzner, D. J., Hall, W., Ohara, K., Shadbolt, N., e Hendler, J. A. (2006). A Framework for Web Science. *Foundations and Trends in Web Science*, 1(1), 1-130. doi:10.1561/1800000001
- Bézivin, J. (1998). Who is afraid of ontologies. *Proceedings of OOPSLA'98 Workshop: " Model Engineering, Methods and Tools Integration with CDIF*.
- Biemann, C. (2005). Ontology Learning from Text: A Survey of Methods. *LDV-Forum 2005 – Band 20 (2)*, pp 75-93.
- Bilhim, J. (2004). *Teoria Organizacional – Estruturas e Pessoas*. Instituto Superior de Ciências Sociais e Políticas Publicações.
- Blue, J., Andoh-Baidoo, F. K., e Osatuyi, B. (2011). An Organizational Memory and Knowledge System (OMKS): Building Modern Decision Support Systems. *International Journal of Data Engineering, IJDE*, 2(2), 27-41.
- Borrajó, M. L., Baruque, B., Corchado, E., Bajo, J., e Corchado, J. M. (2011). Hybrid Neural Intelligent System To Predict Business Failure In Small-To-Medium-Size Enterprises. *International Journal of Neural Systems*, 21(04), 277-296. doi:10.1142/s0129065711002833.
- Bottou, L. (2013). From machine learning to machine reasoning. *Machine Learning*, 94(2), 133-149. doi:10.1007/s10994-013-5335-x.
- Boumarafi, B., e Jabnoun, N. (2008). Knowledge management and performance in UAE business organizations. *Knowledge Management Research & Practice*, 6(3), 233-238. doi:10.1057/kmrp.2008.16.
- Buckinx, W., Verstraeten, G., e Poel, D. V. (2007). Predicting customer loyalty using the internal transactional database. *Expert Systems with Applications*, 32(1), 125-134. doi:10.1016/j.eswa.2005.11.004.
- Buckl, S., Gulden, J., e Schweda, C. (2010). Supporting adhoc Analyses on Enterprise Models. *EMISA. GI, Vol. 172*, pp. 69-83.
- Bumans, G. (2010). Mapping between Relational Databases and OWL Ontologies: an Example. *Scientific Papers, Department of Computing, University of Latvia. Vol. 756*, (pp. 99–117).
- Caiado, A. C. C., e Caiado, J. (2006). *Gestão de Instituições Financeiras*. Edições Sílabo.
- Caldeira, C. (2009). Information Ecology and Domain Definition. *6º CONTECSI - International Conference on Information Systems and Technology Management TECSI/EAC/FEA/USP*.
- Cao, L., Luo, C., Luo, D., e Liu, L. (2004). Ontology Services-Based Information Integration in Mining Telecom Business Intelligence. *PRICAI 2004: Trends in Artificial Intelligence Lecture Notes in Computer Science*, 85-94. doi:10.1007/978-3-540-28633-2_11.
- Cardoso, L. (1998). *Gestão estratégica das organizações: como vencer os desafios do século XXI (5ª Edição)*. Editora Verbo.
- Carvalho, C. (2011). *Simbiose entre DEA e BSC na melhoria do desempenho dos bancos: o caso dos gabinetes de empresas do Banco Alfa*. Tese de doutoramento, Universidade ISCTE-IUL.

- Chapman, A. (2005). Principles of Data Quality, version 1.0. *Report for the Global Biodiversity Information Facility, Copenhagen*. ISBN 87-92020-03-8. Disponível em <http://www.gbif.org/document/80509>, último acesso a 01/04/2015.
- Chaudhuri, S., e Dayal, U. (1997). An overview of data warehousing and OLAP technology. *SIGMOD Record*, 26(1): 65-74.
- Chen, D., Doumeingts, G., e Vernadat, F. (2008). Architectures for enterprise integration and interoperability: Past, present and future. *Computers in Industry*, 59(7), 647-659. doi:10.1016/j.compind.2007.12.016.
- Chen, M., e Plale, B. (2012). From metadata to ontology representation: A case of converting severe weather forecast metadata to an ontology. *Proceedings of the American Society for Information Science and Technology*, 49(1), 1-4. doi:10.1002/meet.14504901286.
- Cheng, H., Lu, Y., e Sheu, C. (2009). An ontology-based business intelligence application in a financial knowledge management system. *Expert Systems with Applications*, 36(2), 3614-3622. doi:10.1016/j.eswa.2008.02.047
- Chiavenato, I. (2005). *Comportamento organizacional: a dinâmica do sucesso das organizações*. São Paulo: Editora Manole.
- Chokri, B. (2007). *Ontology-based Semantic Query Processing in Database Systems*. Doctoral Thesis, Berlin University.
- Chowdhury, T., Tubb, C., e Vidalis, S. (2013). Bridging Semantics Through Ontologies. *SEMANTRO 2013: The Seventh International Conference on Advances in Semantic Processing*.
- Chung, W., Chen, H., e Jr., J. F. (2005). A Visual Framework for Knowledge Discovery on the Web: An Empirical Study of Business Intelligence Exploration. *Journal of Management Information Systems*, 21(4), 57-84. doi:10.1080/07421222.2005.11045821.
- Codd, E. F. (2002). A Relational Model of Data for Large Shared Data Banks. *Software Pioneers*, 263-294. doi:10.1007/978-3-642-59412-0_16.
- Cooper, D. R., Schindler, P. S., e Sun, J. (2003). Business research methods.
- Davenport, T. H., e Prusak, L. (1998). Working knowledge: How organizations manage what they know. *Harvard Business Press*.
- Davis, J. (2010). Information Governance as a Holistic Approach to Managing and Leveraging Information. *BeyeNetwork Custom Research Report Prepared for IBM Corporation*. Disponível em itsmf.cz/wp-content/uploads/2014/04/ibminfegovreport.pdf, último acesso a 01/10/2015.
- Davis, R., Shrobe, H., e Szolovits, P. (1993). What is a knowledge representation?. *AI magazine*, 14(1), 17.
- Dignum, V. (2003). *A model for Organizational Interaction: Based on Agents, Founded in Logic*. Doctoral Thesis, Dutch Research School for Information and Knowledge Systems, SIKS Dissertation Series No. 2004-1.
- Drucker, P. F. (2009). *Managing in a time of great change*. Harvard Business Press.
- Elias, M. (2012). *Enhancing User Interaction with Business Intelligence Dashboards*. Doctoral Thesis, Ecole Centrale Paris: MAS Laboratory.

- Elmasri, R., e Navathe, S. (2003). *Fundamentals of Database Systems* (4a Ed.). Addison-Wesley Longman Publishing Co., Inc.
- Eriksson, H., e Penker, M. (2000). *Business modeling with UML*. New York: OMG Press.
- Farquad, M. A., Ravi, V., e Raju, S. B. (2009). Data Mining Using Rules Extracted from SVM: An Application to Churn Prediction in Bank Credit Cards. *Lecture Notes in Computer Science Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, 390-397. doi:10.1007/978-3-642-10646-0_47
- Fayyad, U., Piatetsky-Shapiro, G., e Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, Volume 17 Number 3.
- Frankel, D., Harmon, P., Mukerji, J., Odell, J., Owen, M., Rivitt, P., Rosen, M., & Soley, R. (2003). *The Zachman Framework and the OMG's Model Driven Architecture*. Disponível em <http://www.bptrends.com/publicationfiles/09-03%20WP%20Mapping%20MDA%20to%20Zachman%20Framework.pdf>, último acesso a 05/09/2015.
- Frosch-Wilke, D., e Tuchtenhagen, S. (2016). Using business intelligence systems for enterprise architecture visualization. *IADIS International Journal on WWW/Internet*, Vol. 14, No. 1, pp. 57-69. ISSN: 1645-7641.
- Gangadharan. G., e Sundaravalli N. (2004). Business Intelligence Systems: Design and Implementation Strategies. *2dh Int. Conf. Information Technology Interfaces IT1 2004*, June 7-10, 2004, Cavtat, Croatia.
- Gartner. (2011). *Gartner Business Analytics Framework*. Disponível em <https://www.gartner.com/doc/1798314/gartners-business-analytics-framework>, último acesso a 05/10/2015.
- Gartner. (2015). *Gartner Magic Quadrant Business Intelligence*. Disponível em <https://www.gartner.com/doc/2989518/magic-quadrant-business-intelligence-analytics>, último acesso a 10/06/2015.
- Gartner. (2015a). *Gartner Magic Quadrant for Enterprise Architecture Tools*. Disponível em <https://www.gartner.com/doc/3162324/magic-quadrant-enterprise-architecture-tools>, último acesso a 01/05/2015.
- Ghawi, R. (2010). *Ontology-based cooperation of information systems: contributions to database-toontology mapping and XML-to-ontology mapping*. Doctoral Thesis, Université de Bourgogne.
- Ghosh, P., Dey, S., e Sengupta, S. (2014). Automatic SQL Query Formation from Natural Language Query. *International Journal of Computer Applications (0975 – 8887), International Conference on Microelectronics, Circuits and Systems (MICRO-2014)*.
- Godinez, M., Koenig, K., Lockwood, S., Oberhofer, M., e Schroeck, M. (2010). *The Art of Enterprise Information Architecture*. Pearson Education India.
- Golfareli, M., e Rizzi, S. (2009). *Data warehouse design: modern principles and Methodologies*. New York: McGraw-Hill Companies.
- Gomez, F., e Rivas, P. (1989). *Estruturas organizativas e informação na empresa*. Editorial Domingos Barreira, 22-407.

Gomez-Perez, A., Fernández-López, M., e Corcho, O. (2006). *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer Science & Business Media.

Gordijn, J., Akkermans, H., e Van Vliet, H. (2000). Business modelling is not process modelling. *ER 2000 Workshops on Conceptual Modeling Approaches for E-Business and The World Wide Web and Conceptual Modeling Salt Lake City, Utah, USA, October 9-12, 2000 Proceedings* (pp. 50-61).

GraphDB. (2016). *Ontotext GraphDB Database*. Disponível em <http://graphdb.ontotext.com/>. Último acesso a 10/06/2015.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199-220. doi:10.1006/knac.1993.1008

Guimaraes, F. (2006). *Utilização de sistemas multi-agentes em processos de negócio e sistemas de workflow no domínio da memória organizacional*. Dissertação de Mestrado, ISCTE.

Hamel, G. (2002). *Leading the revolution: How to thrive in turbulent times by making innovation a way of life*. Boston, MA: Harvard Business School Press.

Hammer, M., e McLeod, D. (1981). Database description with SDM: a semantic database model. *ACM Transactions on Database Systems (TODS)*, 6(3), 351-386.

Handschuh, S. (2005). *Creating Ontology-based Metadata by Annotation for the Semantic Web*. Doctoral Thesis, Fakultät für Wirtschaftswissenschaften der Universität Fridericiana zu Karlsruhe.

Hazman, M., El-Beltagy, S., e Rafea, A. (2011). A Survey of Ontology Learning Approaches. *International Journal of Computer Applications* (0975-8887), Volume 22-No.9, May 2011.

Heflin, J. (2004). *OWL Web Ontology Language-Use Cases and Requirements*. W3C Recommendation, 10, 12.

IBM. (2007). *The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance*. Disponível em https://www-935.ibm.com/services/uk/cio/pdf/leverage_wp_data_gov_council_maturity_model.pdf, último acesso a 01/10/2015.

IBM. (2014). *Banking Data Warehouse Blueprint*. Disponível em <http://www-03.ibm.com/software/products/en/ibm-banking-financial-markets-dw>, último acesso a 15/06/2014.

Ikeda, R., e Widom, J. (2009). *Data lineage: A survey*. Technical report, Stanford University.

Information Management Metamodel (2008). *Information Management Metamodel (IMM) Specification*. OMG document: ad/2008-08-08, disponível em http://www.omgwiki.org/imm/lib/exe/fetch.php?media=submission_i.pdf, último acesso a 10/04/2016.

Inmon, W. H., Strauss, D., e Neushloss, G. (2010). *DW 2.0: the architecture for the next generation of data warehousing*. Morgan Kaufmann.

- Institute of Electrical and Electronics Engineers. (2000). *IEEE Recommended Practice for Architectural Description of Software-Intensive Systems*. IEEE Std 1471-2000, i -23.
- Iyamu, T., e Sekgweleo, T. (2013). Information Systems and Actor-Network Theory Analysis. *International Journal of Actor-Network Theory and Technological Innovation*, 5(3), 1-11. doi:10.4018/jantti.2013070101.
- Jasper, R., e Uschold, M. (1999). A framework for understanding and classifying ontology applications. In *IJCAI 99 Ontology Workshop*, 16–21.
- Johnson, M. W., Christensen, C. M., e Kagermann, H. (2008). Reinventing your business model. *Harvard business review*, 86(12), 57-68.
- Kang, D., Lee, J., Choi, S., e Kim, K. (2010). An ontology-based enterprise architecture. *Expert Systems with Applications* 37 (2010), 1456–1464.
- Kaplan, R. S., e Norton, D. P. (2008). Mastering the management system. *Harvard business review*, 86(1), 62.
- Kaur, S., e Bali, R. (2012). Sql Generation Execution from Natural Language Processing. *International Journal of Computing & Business Research ISSN (Online): 2229-6166*
- Keim, D., Kohlhammer, J., Ellis, G., e Mansmann, F. (2010). *Solving problems with visual analytics*. Goslar: Eurographic Association.
- Kennerley, M., e Neely, A. (2003). Measuring Performance in a Changing Business Environment. *International Journal of Operations & Production Management*, 23, 2, 213-229.
- Keri, E., Pearlson, S., e Saunders, C. (2006). *Managing and using information systems: A strategic approach*. USA: Wiley.
- Khalid, S. (2010). *Towards Data Governance for International Dementia Care Mapping (DCM): A Study Proposing DCM Data Management through a Data Warehousing Approach*. Doctoral Thesis, School of Computing, Informatics and Media University of Bradford. Disponível em <https://bradscholars.brad.ac.uk/handle/10454/5226>, ultimo acesso a 01/01/2015.
- Kimbal, R., e Ross, M. (2013). *The Data Warehouse Toolkit: The definitive guide to dimensional model (3a Ed.)*. USA: Wiley.
- Kotler, P., e Keller, K. (2006). *Marketing Management (12a Ed.)*. New Jersey: Pearson Prentice-Hall.
- Kremen, P. (2012). *Building Ontology-Based Information Systems*. Doctoral Thesis, Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Cybernetics.
- Krummenacher, R., e Strang, T. (2007). Ontology-Based Context Modeling. *Workshop on Context-Aware Proactive Systems Issue 2007*.
- Kueri (2016). *Natural Language to SQL Interface for Relational Database*. Disponível em <http://kueri.me/>, último acesso a 10/05/2016.
- Kung, P., Hagen, C., Rodel, M., e Seifert, S. (2005). Business process monitoring & measurement in a large bank: challenges and selected approaches. *16th International Workshop on Database and Expert Systems Applications (DEXA'05)* (pp. 955-961), IEEE.

- Kuo, Y., Lonie, A., Sonenberg, L., e Paizis, K. (2007). Domain ontology driven data mining: A medical case study. *Domain Driven Data Mining DDDM2007*, A. S. W., editor, *Conference on Knowledge Discovery in Data* (pp. 11–17), California USA, ACM.
- Land, M., Proper, E., Waage, M., Cloo, J., e Steghuis, C. (2009). *Enterprise Architecture - Creating Value by Informed Governance*. Springer. DOI: 10.1007/978-3-540-85232-2.
- Lankhorst, M. (2006). *Enterprise Architecture at Work*. Springer.
- Laudon, K., e Laudon, J. (2000). *Management information system: organization and technology in the networked enterprise*. Pearson Custom Publishing.
- Laudon, K., e Laudon, J. (2012). *Sistemas de Información Gerencial*. Pearson.
- Law, J. (1992). *Notes on the Theory of the Actor Network: Ordering, Strategy and Heterogeneity*. Centre for Science Studies, Lancaster University, Lancaster LA1 4YN. Disponível em <http://www.comp.lancs.ac.uk/sociology/papers/Law-Notes-on-ANT.pdf>, ultimo acesso a 01/06/2015.
- Lependu, P. (2010). *Ontology databases*. Doctoral Thesis, Graduate School of the University of Oregon.
- Liles, D. H., Johnson, M. E., Meade, L., & Underdown, D. R. (1995). Enterprise Engineering: A Discipline?. *Society for Enterprise Engineering Conference Proceedings*, Vol. 6, pp. 45-47.
- Locke, K. (2001). *Grounded Theory in Management Research*. SAGE Publications Ltd.
- Lubyte, L. (2010). *Techniques and Tools for the Design of Ontologies for Data Access*. Doctoral Thesis, KRDB Research Centre, Faculty of Computer Science, Italy.
- Maedche, A., e Staab, S. (2001). Ontology learning for the Semantic Web. *IEEE Intelligent Systems and Their Applications* 16(2): 72-79.
- Maise, M. (2013). *Connecting Enterprise Architecture and Information Objects Using an Enterprise Ontology*. Master Thesis, University of Applied Sciences and Arts, Northwestern Switzerland School of Business.
- Marco, D. (2013). *Meta Data Silos*. Disponível em <http://tdan.com/meta-data-silos/5171>, último acesso a 12/08/2015.
- Martinez, M. T., Fouletier, P., Park, K. H., e Favrel, J. (2001). *Virtual enterprise—organisation, evolution and control*. *International journal of production economics*, 74(1), 225-238.
- Mawilmada, P. K. (2011). *Impact of a data warehouse model for improved decision-making process in healthcare*. Master Thesis, Faculty of Science and Technology, Queensland University of Technology.
- Mendialdua, Z. (2011). *Harnessing Folksonomies for resource classification*. Doctoral Thesis, Universidad Nacional de Educación a Distância Departamento de Lenguajes y Sistemas Informáticos, Escuela Técnica Superior de Ingeniería Informática.
- Microsoft English Query (2016). *Microsoft English Query Authoring Tool*. Disponível em <https://technet.microsoft.com/en-us/library/cc966482.aspx>, último acesso a 10/06/2016.
- Microsoft SQLServer Metadata Toolkit. (2016). Disponível em <https://sqlmetadata.codeplex.com/>, último acesso a 10/06/2015.

- Microsoft SQLServer Semantic DB. (2014). *SQLServer Semantic DB*. Disponível em <https://msdn.microsoft.com/en-us/library/gg509085.aspx>, último acesso a 05/05/2014.
- Microsoft. (2009). *Microsoft Enterprise Architecture*. Disponível em <https://msdn.microsoft.com/en-us/library/aa480017.aspx>, último acesso a 05/09/2015.
- Mochól, M. (2009). *The Methodology for Finding Suitable Ontology Matching Approaches*. Doctoral Thesis, Institut für Informatik des Fachbereichs Mathematik und Informatik der Freien Universität Berlin.
- Mohania, M., Samtami, S., Roddick, J., e Kambayashi, Y. (1999). Advances and research directions in data-warehousing technology. *Australasian Journal of Information Systems*, Vol 7, No 1.
- Morais, A. (2013). *Aplicação de Ontologias à Representação e Análise de Modelos de Arquitetura Empresarial*. Dissertação de Mestrado, Instituto Superior Técnico, Lisboa.
- Morais, E., e Ambrósio, A. (2007). *Ontologias: conceitos, usos, tipos, metodologias, ferramentas e linguagens*. Relatório Técnico, INF_001/07, Instituto de Informática, Universidade Federal de Goiás.
- Morris, M., Schindehutte, M., e Allen, J. (2005). The entrepreneur's business model: toward a unified perspective. *Journal of business research*, 58(6), 726-735.
- Moss, L., e Atre, S. (2003). *Business intelligence roadmap: the complete project lifecycle for decision-support applications*. Addison-Wesley Professional.
- Mote, K. (2002). *Natural Language Processing: A Survey*. Computing Research Repository, Cpts499, May 2002.
- Myers, M. D. (1997). Qualitative research in information systems. *Management Information Systems Quarterly*, 21(2), 241-242.
- Nahm, Y., e Ishikawa, H. (2005). A hybrid multi-agent system architecture for enterprise integration using computer networks. *Robotics and Computer-Integrated Manufacturing*, 217-234.
- National Information Standards Organization. (2004). *Understanding Metadata*. Disponível em <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>, último acesso a 02/05/2015.
- Neely, A. (2002). *Avaliação do Desempenho das Empresas*. Editorial Caminho.
- Neely, A., Gregory, M., e Platts, K. (2005). Performance Measurement System Design – a Literature Review and Research Agenda. *International Journal of Operations & Production Management*, 25, 12, 1228-1263.
- Nemati, H., Steiger, D., Iyer, L., e Herschel, R. (2002). Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and datawarehousing. *Decision Support Systems*, 33, 143–161.
- Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *The strategic management of intellectual capital and organizational knowledge (2002)*, 437-462.

- Noy, N., e McGuinness, D. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. Technical Report KSL-01-05, Palo Alto, California, USA: Knowledge Systems Laboratory, Stanford University.
- O'Brien, J. (2004). *Sistemas de informação e as decisões gerenciais na era da internet*. São Paulo: Saraiva.
- Ochoa, S. F., Herskovic, V., Pineda, E., e Pino, J. A. (2009). A transformational model for Organizational Memory Systems management with privacy concerns. *Information Sciences*, 179(15), 2643-2655.
- Orofino, M. A. R. (2011). *Técnicas de criação do conhecimento no desenvolvimento de modelos de negócio*. Dissertação de Mestrado, Universidade Federal de Santa Catarina.
- Ostaszadeh, S., Aliee, F., e Ostaszadeh, A. (2003). A Method for Consistent Modeling of Zachman Framework Cells. In K. Elleithy (Ed.), *Advances and Innovations in Systems, Computing Sciences and Software Engineering*, The Netherlands: Springer, pp. 375-380.
- Osterwalder, A. (2004). *The Business Model Ontology. A Proposition in a Design Science Approach*. Doctoral Thesis, Université de Lausanne, Ecole des Hautes Etudes Commerciales.
- Osterwalder, A., e Pigneur, Yves. (2010). *Criar Modelos de Negócio*. Publicações Dom Quixote.
- Osterwalder, A., Pigneur, Y., e Tucci, C. L. (2005). Clarifying Business Models: Origins, Present and Future of the Concept. *Communications of the Association for Information Science (CAIS)*, 16: 1-25.
- Othero, G. (2006). Linguística Computacional: uma breve introdução. *Letras de hoje*, 41(2), 341-351.
- Panetto, H., e Molina, A. (2008). Enterprise integration and interoperability in manufacturing systems: Trends and issues. *Computers in Industry*, 641-646
- Pereira, D. (2012). *Engenharia de ontologias para redes colaborativas*. Dissertação de Doutoramento, Faculdade de Engenharia, Universidade do Porto.
- Perrin, O., e Godart, C. (2004). A model to support collaborative work in virtual enterprises. *Data & Knowledge Engineering*, 50(1), 63-86.
- Piatesky-Shapiro, G., e Matheus, C. (1992). Knowledge discovery workbench for exploring business databases. *International Journal of Intelligent Agents*, 7(7):675-686
- Pinto, F. (2009). *Database Marketing Intelligence Methodology Supported by Ontologies and Knowledge Discovery in Databases*. Doctoral Thesis, Universidade do Minho.
- PlanView. (2017). PlanView Enterprise Architecture Solution. Disponível em <http://www.planview.com/tapm-technology-application-portfolio-management/>, último acesso a 19/03/2017.
- Porter, M. (1985). *Competitive advantage: creating and sustaining superior performance*. 1985. New York: FreePress.
- Potter, R., Rainer, K., e Turban, E. (2005). *Administração de Tecnologia da Informação – Teoria e Prática* (3a Ed.). Rio de Janeiro: Campus.

- Rajabi, Z., Minaei, B., e Seyyedi, M. (2013). Enterprise Architecture Development Based on Enterprise Ontology. *Journal of Theoretical and Applied Electronic Commerce Research*, Electronic Version VOL 8 / ISSUE 2 / AUGUST 2013 / 85-95. ISSN 0718-1876,
- Ramos, S., e Naranjo, E. (2013). *Metodologia da investigação científica*. Escolar Editora.
- Rao, M. (2012). *Knowledge management tools and techniques*. New York: Routledge.
- Rittgen, P. (2006). *Enterprise modeling and computing with UML*. Idea Group Publishing.
- Robson, C. (1993). *Real world research: A resource for social scientists and practitioners-researchers*. Massachusetts: Blackwell Publishing.
- Rodrigues, L. (2014). *Fatores determinantes do valor das arquiteturas empresariais*. Tese de doutoramento, Universidade do Minho, Escolha de Engenharia.
- Rodriguez-Muro, M., Lubyte, L., e Calvanese, D. (2008, April). Realizing ontology based data access: A plug-in for Protégé. *Data Engineering Workshop 2008, ICDEW 2008, IEEE 24th International Conference* (pp. 286-289).
- Rosini, A., e Palmisano, A. (2003). *Administração de sistemas de informação e a gestão do conhecimento*. São Paulo: Pioneira Thomson Learning.
- Russo, J. (2006). *Balanced scorecard para PME*. Lidel—edições técnicas.
- Saias, J., Quaresma, P., Salgueiro, P., e Santos, T. (2012). BINLI: An Ontology-Based Natural Language Interface for Multidimensional Data Analysis. *Intelligent Information Management*, 2012, 4, 225-230, Published Online September 2012. doi:10.4236/iim.2012.45033.
- Santos, S. P., Belton, V., e Howick, S. (2002). Adding value to performance measurement by using system dynamics and multicriteria analysis. *International Journal of Operations & Production Management*, 22(11), 1246-1272.
- Saunders, M., Lewis, P., e Thornhill, A. (1987). *Research Methods for Business Students* (4a Ed.). Pearson Education Limited.
- Sciarrone, F., Starace, P., e Federici, T. (2009). A business intelligence process to support information retrieval in an ontology-based environment. *2009 Ninth International Conference on Intelligent Systems Design and Applications* (pp. 896-901).
- Shafer, S. M., Smith, H. J., e Linder, J. C. (2005). The power of business models. *Business horizons*, 48(3), 199-207.
- Shan, W., e Zhang, Q. (2007). Study on Knowledge Mining of the Business Intelligence System. *2007 International Conference on Wireless Communications, Networking and Mobile Computing*. doi:10.1109/wicom.2007.1331
- Sicilia, M. (2006). A. Metadata, semantics, and ontology: providing meaning to information resources. *International Journal of Metadata, Semantics and Ontologies*, Vol. 1, No. 1.
- Simard, C., e Rice, R. (1996). The Practice Gap: Barriers to the Diffusion of Best Practices. In C. R. McInerney & R. E. Day (Eds.), *Rethinking Knowledge Management: From Knowledge Objects to Knowledge Processes* (pp. 87-124). New York, USA. Springer.

- Simghan, Y., Plale, B., e Gannon, D. (2005). *A Survey of Data Provenance Techniques*. Technical Report TR-618: Computer Science Department, Indiana University.
- Singh, S. (2013). An Experiment in Software Component Retrieval based on Metadata and Ontology Repository. *International Journal of Computer Applications* (0975 – 8887), Volume 61, No.14, January 2013.
- Skyrme, D. J., e Amidon, D. M. (1998). New measures of success. *Journal of Business Strategy*, 19(1), 20-24.
- Smith, H., e Fingar, P. (2003). *Business process management: the third wave* (Vol. 1). Tampa: Meghan-Kiffer Press.
- Sowa, J. (2000a). *Process and Causality*. Disponível em <http://www.ifsowa.com/ontology/causal.htm>, último acesso a 01/06/2016.
- Sowa, J. (2000b). Ontology, metadata na semiotics. *Lecture Notes in Computer Science Conceptual Structures: Logical, Linguistic, and Computational Issues*, 55-81. doi:10.1007/10722280_5
- Stefanov, V. (2007). *Conceptual Models and Model-Based Business Metadata to Bridge the Gap between Data Warehouses and Organizations*. Doctoral Thesis, Vienna University of Technology, Institute of Software Technology and Interactive Systems.
- Stowell, F., West, D., e Stansfield, M. (1997). Action research as a framework for IS research. *Information Systems: an emerging discipline*, 159-200.
- Szwed, P. (2015). *ArchiOWLPlug-In*. Disponível em http://home.agh.edu.pl/~pszwed/en/doku.php?id=archi_to_owl, último acesso a 20/01/2015
- Szwed, P., Komnata, W., e Dymek, D. (2015). DWARM: An Ontology of Data Warehouse Architecture Reference Model. Beyond Databases, Architectures and Structures. *Communications in Computer and Information Science*, Volume 521, 2015, pp 222-232.
- Tannenbaum, A., e Foreword By-Simon, A. (2001). *Metadata solutions: using metamodels, repositories, XML, and enterprise portals to generate information on demand*. Addison-Wesley Longman Publishing Co., Inc.
- Teece, D. J. (2010). Business models, business strategy and innovation. *Long range planning*, 43(2), 172-194.
- Telea, A. (2007). *Data Visualization: Principles and practices (2a Ed.)*. Boca Raton: CRC Press.
- Teradata. (2014). *Teradata Financial Services Data Warehouse*. Disponível em <http://www.teradata.com/logical-data-models/financial-services/>, último acesso a 15/06/2014.
- The Open Group Architecture Framework. (2013). *The Open Group Architecture Framework*. Disponível em www.Togaf.org, último acesso 01/05/2015.
- The Open Group. (2012). *ArchiMate 2.0 Specification*. Disponível em <http://pubs.opengroup.org/architecture/ArchiMate2-doc/>, último acesso a 10/03/2015.
- The Open Group. (2016). *ArchiMate 3.0.1 Specification*. Disponível em <http://pubs.opengroup.org/architecture/archimate3-doc/>, último acesso a 10/05/2016.

- Tiwana, A. (2003). Knowledge partitioning in outsourced software development: A field study. *ICIS 2003 Proceedings*, 22.
- Towers, S. (2005). *In Search of BPM Excellence: Straight From The Thought Leaders*. Meghan-Kiffer Press.
- Tozer, G. (1999). *Metadata management for information control and business success*. Artech House, Inc.
- Tsichritzis, D., e Lochovsky, F. (1976). Hierarchical data-base management: A survey. *ACM Computing Surveys (CSUR)*, 8(1), 105-123.
- Turco, C. (2010). *Enterprise Architecture & Metadados Modelling: A Guide to Conceptual Data Model, Metadados Repositoty, Business and System Re-Enginerring*. Infinity Publishing. ISBN-0-7414-5304-5.
- Van Der Aalst, W. (2013). Process Cubes: Slicing, Dicing, Rolling Up and Drilling Down Event Data for Process Mining. *Lecture Notes in Business Information Processing Asia Pacific Business Process Management*, 1-22. doi:10.1007/978-3-319-02922-1_1
- Vand Den Berg, G., e Pietersma, P. (2014). *Os principais modelos de gestão: 77 modelos que todo o gestor devia conhecer*. Actual. ISBN: 978-989-694-125-3.
- Vetterli, T., Vaduva, A., e Staudt, M. (2000). Metadata standards for data warehousing: open information model vs. common warehouse metadata. *ACM Sigmod Record*, 29(3), 68-75.
- Villanueva-Rosales, N. (2011). *Formalizing Relational Databases as OWL Ontologies*. Doctoral Thesis, Carleton University Canada.
- Villiers, M. (2005). Three approaches as pillars for interpretive information systems research: development research, action research and grounded theory. *SAICSIT - 2005 annual research conference of the South African institute of computer scientists and information technologists*, White River.
- Walsh, J., e Ungson, G. (1991). Organizational memory. *Academy of management review*, 16(1), 57-91.
- Webjornsen, R. (2005). *Discovering data lineage in Datawarehouse: Methods and techniques for tracing the origins of data in data-warehouse*. Master Thesis, University of Oslo, Norway.
- Wei, S., e Qing-pu, Z. (2007). Study on knowledge mining of the business intelligence system. *WICOM 2007, International Conference on Wireless communications, networking and mobile computing*, pp. 5435-5438.
- Whitman, L., e Panetto, H. (2006). The missing link: Culture and language barriers to interoperability. *Annual Reviews in Control*, 233-241.
- Wickett, K. (2012). *Collection/Item Metadata Relationships*. Doctoral Thesis, Graduate College of the University of Illinois at Urbana-Champaign.
- Wizdee. (2016). *Natural Language BI*. Disponível em <http://wizdee.com/home/>, último acesso a 01/06/2016.
- Won, K. (2005). On Metadata Management Technology: Status and Issues. *Journal of Object Technology*, Vol 4, No 2, March-April 2005.

World Wide Web Consortium. (2014). *SPARQL Protocol and RDF Query Language*. Disponível em <http://www.w3.org/TR/sparql11-overview/>, último acesso a 19/03/2014.

World Wide Web Consortium. (2013). *OWL Web Ontology Language*. Disponível em <http://www.w3.org/TR/owl2-syntax/>, último acesso a 12/08/2013.

Yin, R. K. (2013). *Case study research: Design and methods*. SAGE publications Ltd.

Zachman, J. (2009). *The Zachman Framework Evolution*. Disponível em <https://www.zachman.com/ea-articles-reference/54-the-zachman-framework-evolution>, último acesso a 01/05/2015.

Zhou, S., Ling, H., Han, M., e Zhang, H. (2010). Ontology Generator from Relational Database Based on Jena. *Computer and Information Science*, 3(2). doi:10.5539/cis.v3n2p263

Anexo A - Questionário a empresas sobre AE e BI

Relação entre Arquitectura Empresarial (AE) e Modelo Informacional de BI

Este questionário enquadra-se numa tese de Doutoramento em Enterprise Intelligence de Francisco Guimarães da Universidade de Évora sob orientação do Dr. Carlos Pampulim e Dr. Paulo Quaresma.

A investigação está focada no problema de manter uma integração entre Arquitecturas Empresariais (AE) e Business Intelligence (BI) tendo por base a arquitetura de informação e metadados corporativos. Para o efeito recorre a ontologias em OWL para criar uma interoperabilidade entre AE e BI, no sentido de geração automática de AE a partir de BI (base de dados e classificação semântica de métricas e dimensões) como teste de hipótese de solução num dos sentidos da interoperabilidade.

O questionário tem por objetivo obter uma perceção sobre o nível de utilização de arquitecturas empresariais e sua possível relação com a arquitetura de BI considerando uma estrutura de metadata corporativa de conceitos e relações entre conceitos presentes nas duas arquitecturas, enquanto tema em investigação.

*Required

1. Qual a entidade onde trabalha?

.....

2. Qual a sua função e unidade orgânica?

.....

Arquitectura Empresarial

Esta seção tem por objetivo capturar informação base sobre o nível e tipo de utilização da arquitectura empresarial.

3. **A quantos anos existe a Arquitectura Empresarial ***

Tick all that apply.

- A mais de 5 anos
- Entre 5 e 3 anos
- Entre 3 e 1 ano
- Other:

4. **Que Unidades Orgânicas estão envolvidas na gestão da AE? ***

Tick all that apply.

- Planeamento Estratégico
- Organização e Métodos
- Sistemas de Informação
- Other:

5. **Que Metamodelo é utilizado para AE? ***

Tick all that apply.

- Archimate
- Zachman
- Proprietário da ferramenta
- Other:

6. **Que níveis de arquitetura de AE existem no seu modelo? ***

Tick all that apply.

- Negócio sem BSC
- Negócio com BSC
- Orgânica
- Processos ou Funcional
- Aplicacional
- Informacional
- Tecnológica
- Other:

7. **Que ferramenta é utilizada para AE?**

.....

8. **A ferramenta tem integração com outras fontes de artefactos? ***
Tick all that apply.

- Sistemas de BPM/Workflow
- CMDB (Configuration Management DB)
- SOA
- Metadados BI
- Gestão de Capital Humano
- Catálogo de Produtos/Serviços
- Other:

9. **Qual o nível de dificuldade na criação da AE? ***
Mark only one oval.

	1	2	3	
Elevado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Baixo

10. **Qual o nível de dificuldade na atualização da AE?**
Mark only one oval.

	1	2	3	
Elevado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Baixo

11. **Qual o principal motivo na dificuldade de gestão da AE? ***
Tick all that apply.

- Envolvimento da organização
- Obtenção de informação de ativos da organização
- Obtenção de informação de gestão para KPI's
- Alterações constantes de negócio
- Other:

12.

Qual a utilização da AE na Organização? *

Tick all that apply.

- Em fase de planeamento estratégico de negócio
- Em fase de planeamento estratégico de SI
- Na análise de impacto de alterações em SI
- Na documentação de processos de negócio
- Na análise de risco de sistemas de informação
- Other:

Arquitectura de BI

Esta seção tem por objetivo entender a utilização de metadados e níveis de arquitetura de Business Intelligence.

13.

A quantos existe a arquitetura de BI? *

Mark only one oval.

- A mais de 5 anos
- Entre 5 e 3 anos
- Entre 3 e 1 ano
- Other:

14.

Que Unidades Orgânicas gerem a arquitetura de BI? *

Tick all that apply.

- Unidade específica de BI
- Unidade de gestão de aplicações de SI
- Marketing
- Planeamento Estratégico
- Other:

15.

Que (meta)modelo é utilizado para BI? *

Mark only one oval.

- Desenhado pela organização
- Adaptação de modelo do fabricante da ferramenta
- Other:

16. **Que níveis de arquitetura de BI existem? ***

Tick all that apply.

- ODS
- DataWareHouse
- DataMart
- Reporting Fixo
- Reporting em Self-Service
- Dashboard
- BigData Analytics
- Data Mining
- Function Analytics (Profitability, Customer Intelligence, Risk Intelligence, outros)
- Other:

17. **Que tipo de metadata existe? ***

Tick all that apply.

- Não existe
- Gerada e utilizada no ETL
- Gerada e utilizada pelas ferramentas de exploração de BI
- Metadata transversal para ETL e Exploração BI
- Metadata Corporativa para BI e sistemas operacionais
- Other:

18. **Que ferramentas são utilizadas para BI (ETL, Metadata e Exploração BI)?**

.....

19. **Qual o nível de dificuldade na gestão de BI? ***

Mark only one oval.

	1	2	3	
Elevado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Baixo

20.

Qual o principal motivo de dificuldade de gestão de BI? *

Tick all that apply.

- Velocidade de necessidades de utilizadores para novos Dashboard
- Velocidade de necessidades de utilizadores para novos Reporte's
- Captura de novas métricas e dimensões a partir de sistemas operacionais
- Alterações de sistemas operacionais sem acompanhamento pelo BI
- Manutenção de metadados consistentes entre conceitos de negócio e conceitos de BI
- Other:

Teste de hipótese de relação entre AE e BI

Esta seção tem por objetivo analisar a importância de uma potencial relação entre os modelos de AE e BI.

21.

Qual o nível de importância do modelo AE como ontologia corporativa? *

Tick all that apply.

- Não é relevante
- Relevante porque contém todos os conceitos e relações (semântica) do negócio
- Relevante mas deve ser complementado com glossários de termos
- Relevante mas deve ser complementado com modelos de dados (tabelas, campos e relações)
- Other:

22.

Qual o nível de importância de classificar e reutilizar dimensões em BI a partir de conceitos e relações em AE? *

Tick all that apply.

- Não é relevante
- Não é possível porque em AE não existem todas os conceitos e relações correspondentes as dimensões em BI
- Otimizaria o esforço de modelação de dimensões e hierarquias em BI
- Facilitaria a gestão corporativa de metadados entre AE e BI
- Other:

23.

Qual o nível de importância de classificar e reutilizar métricas em BI a partir de KPI e objetivos em AE? *

Tick all that apply.

- Não é relevante
- Não é possível pois em AE não se consegue ter todas as formulas de cálculo como regras de negócio
- Otimizaria o esforço de relacionar KPI em AE com métricas em BI
- Other:

24.

Qual o nível de importância de Linguagem Natural como interface com utilizador em AE e BI? *

Tick all that apply.

- Não é relevante
- Otimiza a forma de pesquisa para utilizadores menos "tecnológicos"
- Other:

Powered by



Anexo B – Questionário casos estudo

Relação entre Arquitetura Empresarial (AE) e Business Intelligence (BI)

Este questionário enquadra-se numa tese de Doutoramento em Enterprise Intelligence de Francisco Guimarães da Universidade de Évora sob orientação do Dr. Carlos Pampulim e Dr. Paulo Quaresma.

A investigação está focada no problema de manter uma integração entre Arquiteturas Empresariais (AE) e Business Intelligence (BI) tendo por base a arquitetura de informação e metadados corporativos. Para o efeito recorre a ontologias em OWL para criar uma interoperabilidade entre AE e BI, no sentido de geração automática de AE a partir de BI (base de dados e classificação semântica de métricas e dimensões) como teste de hipótese de solução num dos sentidos da interoperabilidade.

O questionário tem por objetivo obter uma perceção sobre o nível de utilização de arquiteturas empresariais e sua possível relação com a arquitetura de BI considerando uma estrutura de metadata corporativa de conceitos e relações entre conceitos presentes nas duas arquiteturas, enquanto tema em investigação.

***Required**

1. Qual a entidade onde trabalha?

.....

2. Qual a sua função e unidade orgânica?

.....

Arquitetura empresarial ou documentação equivalente

Esta seção tem por objetivo entender o tipo de documentação sobre arquiteturas que existe atualmente na organização.

3. **Porque razão não existe Arquitetura Empresarial (AE) ***

Tick all that apply.

- Documentação de processos é suficiente
- Documentação de SI é suficiente
- Organograma é suficiente
- Plano estratégico é suficiente
- Other:

4. **Que Unidades Orgânicas deviam estar envolvidas na gestão da AE ou documentação equivalente? ***

Tick all that apply.

- Planeamento Estratégico
- Organização e Métodos
- Sistemas de Informação
- Other:

5. **Que níveis de arquitetura deveriam existir no seu modelo em AE ou documentação equivalente? ***

Tick all that apply.

- Negócio sem BSC
- Negócio com BSC
- Orgânica
- Processos ou Funcional
- Apicacional
- Informacional
- Tecnológica
- Other:

6. **Que ferramenta é utilizada para documentar arquiteturas da empresa?**

.....

7. **A ferramenta tem integração com outras fontes? ***

Tick all that apply.

- Sistemas de BPM/Workflow
- CMDB (Configuration Management DB)
- SOA
- Metadados BI
- Gestão de Capital Humano
- Catálogo de Produtos/Serviços
- Other:

8. **Qual o nível de dificuldade na atualização da documentação?**
Mark only one oval.

	1	2	3	
Elevado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Baixo

9. **Qual o principal motivo na dificuldade de documentação? ***
Tick all that apply.

- Envolvimento da organização
- Obtenção de informação de ativos da organização
- Obtenção de informação de gestão para KPI's
- Alterações constantes de negócio
- Alinhamento de conceitos entre negócio e SI
- Other:

10. **Que utilização faz da documentação de arquiteturas? ***
Tick all that apply.

- Em fase de planeamento estratégico de negócio
- Em fase de planeamento estratégico de SI
- Na análise de impacto de alterações em SI
- Na documentação de processos de negócio
- Na análise de risco de sistemas de informação
- Other:

Arquitetura de BI

Esta seção tem por objetivo entender a utilização de metadados e níveis de arquitetura de Business Intelligence existente.

11.

Que níveis de arquitetura de BI existem? *

Tick all that apply.

- ODS
- DataWareHouse
- DataMart
- Reporting Fixo
- Reporting em Self-Service
- Dashboard
- BigData Analytics
- Data Mining
- Function Analytics (Profitability, Customer Intelligence, Risk Intelligence, outros)
- Other:

12.

Que Unidades Orgânicas gerem a arquitetura de BI? *

Tick all that apply.

- Unidade específica de BI
- Unidade de gestão de aplicações de SI
- Marketing
- Planeamento Estratégico
- Other:

13.

Que (meta)modelo é utilizado para BI? *

Mark only one oval.

- Desenhado pela organização
- Adaptação de modelo do fabricante da ferramenta
- Other:

14.

Que tipo de metadata existe? *

Tick all that apply.

- Não existe
- Gerada e utilizada no ETL
- Gerada e utilizada pelas ferramentas de exploração de BI
- Metadata transversal para ETL e Exploração BI
- Metadata Corporativa para BI e sistemas operacionais
- Other:

15. **Que ferramentas são utilizadas para BI (ETL, Metadata e Exploração BI)?**

.....

16. **Qual o nível de dificuldade na gestão de BI? ***
Mark only one oval.

1 2 3

Elevado Baixo

17. **Qual o principal motivo de dificuldade de gestão de BI? ***
Tick all that apply.

- Velocidade de necessidades de utilizadores para novos Dashboard
- Velocidade de necessidades de utilizadores para novos Reporte's
- Captura de novas métricas e dimensões a partir de sistemas operacionais
- Alterações de sistemas operacionais sem acompanhamento pelo BI
- Manutenção de metadados consistentes entre conceitos de negócio e conceitos de BI
- Other:

Teste de hipótese de relação entre AE e BI

Esta seção tem por objetivo analisar a importância de uma potencial relação entre os modelos de AE e BI.

18. **Qual a sua opinião sobre as causas e impacto do problema indicado no capítulo 2.3 ***

Escolher duas das três hipóteses
Mark only one oval.

- Concordo
- Não concordo
- Other:

19. **Qual a sua opinião sobre utilizar a arquitetura empresarial como base para definir métricas e dimensões em BI? ***

Escolher duas das três hipóteses
Mark only one oval.

- Concordo
- Não concordo
- Other:

20. Qual a sua opinião sobre o resultado do inquérito no capítulo 4.2 face à sua organização? *

Mark only one oval.

- Concordo
- Não concordo
- Other:

21. A arquitetura organizacional apresentada reflete a realidade da sua organização? *

Mark only one oval.

1	2	3	4	5
<input type="radio"/>				

22. A arquitetura de negócio apresentada reflete a realidade da sua organização? *

Mark only one oval.

	1	2	3	4	5	
Não	<input type="radio"/>	Sim				

23. A base de dados apresentada reflete as dimensões e factos críticos da sua organização? *

Mark only one oval.

	1	2	3	4	5	
Não	<input type="radio"/>	Sim				

24. Os Dashboard apresentados refletem a estrutura de análises (sem a totalidade das métricas) críticas da sua organização alinhada com a arquitetura empresarial?

Mark only one oval.

	1	2	3	4	5	
Não	<input type="radio"/>	Sim				

25.

O analisador de expressões apresentado corresponde a uma linguagem de interrogação de dimensões e métricas em BI? *

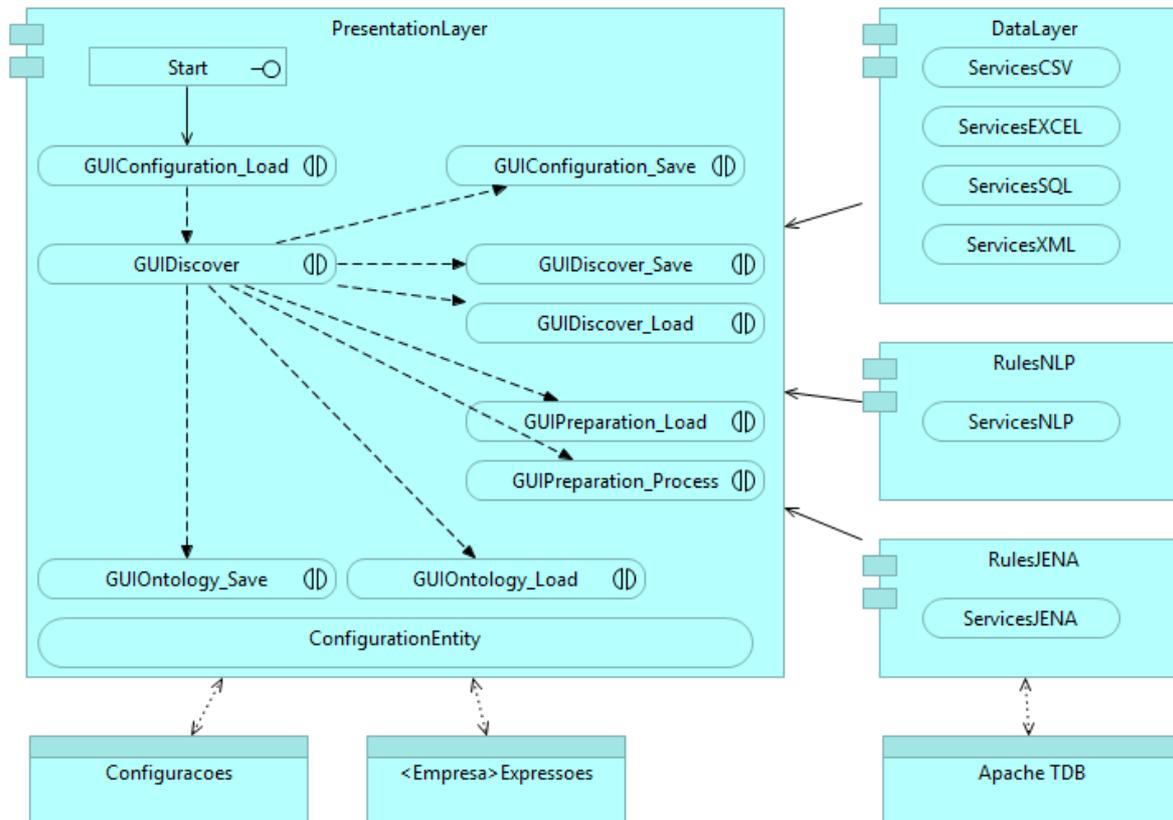
Mark only one oval.

	1	2	3	4	5	
Não	<input type="radio"/>	Sim				

Powered by
 Google Forms

Anexo C – Artefatos do protótipo em Java

O código Java é composto por classes e métodos integrados em Packages de acordo com a figura seguinte, estando disponíveis para visualização no *website* disponibilizado para o efeito em <https://eaintelligence.wordpress.com>.





UNIVERSIDADE DE ÉVORA
INSTITUTO DE INVESTIGAÇÃO
E FORMAÇÃO AVANÇADA

Contactos:

Universidade de Évora

Instituto de Investigação e Formação Avançada - IIFA

Palácio do Vimioso | Largo Marquês de Marialva, Apart. 94

7002-554 Évora | Portugal

Tel: (+351) 266 706 581

Fax: (+351) 266 744 677

email: iifa@uevora.pt