



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Abordagem Ontológica na Gestão de Dados: Estudo de Caso CGU

Márcia Myuki Takenaka Fujimoto

Dissertação apresentada como requisito parcial para conclusão do
Mestrado Profissional em Computação Aplicada

Orientadora
Prof.a Dr.a Edna Dias Canedo

Brasília
2018

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

TM319a TAKENAKA FUJIMOTO, MÁRCIA MYUKI
ABORDAGEM ONTOLÓGICA NA GESTÃO DE DADOS: ESTUDO DE CASO
CGU / MÁRCIA MYUKI TAKENAKA FUJIMOTO; orientador EDNA DIAS
CANEDO. -- Brasília, 2018.
77 p.

Dissertação (Mestrado - Mestrado Profissional em
Computação Aplicada) -- Universidade de Brasília, 2018.

1. GESTÃO DE DADOS. 2. ONTOLOGIA. 3. INTEGRAÇÃO DE
DADOS. 4. ARQUITETURA DA INFORMAÇÃO. I. DIAS CANEDO, EDNA,
orient. II. Título.

Dedicatória

Dedico esse trabalho aos meus pais Tamotsu e Keoko, meus referenciais de coragem e persistência. Aos meus filhos, Bárbara e Vítor, e ao meu companheiro de vida, Nelson, agradeço cada minuto de paciência e apoio, imprescindíveis à realização deste trabalho e de muitos outros no campo profissional.

Agradecimentos

Agradeço à Prof.^a Dr.^a Edna Dias Canedo, minha orientadora, pelo suporte essencial durante este trabalho.

Aos professores e colegas do Mestrado Profissional em Computação Aplicada por todo o conhecimento compartilhado.

Aos colegas do Ministério da Transparência e Controladoria-Geral da União pelo constante auxílio e compreensão durante o mestrado.

Resumo

Este trabalho estuda a utilização da ontologia como abordagem para a gestão e integração de dados, ela é considerada como meio de dominar a complexidade envolvida em ambientes com grande diversidade de fontes de dados. A aplicabilidade variada da ontologia direcionou as investigações deste trabalho no sentido de considerá-la como base de solução apresentada para o estudo de caso considerado: arquitetura da informação para a gestão de dados CGU. A CGU é o Ministério do poder executivo federal com responsabilidade primária pelas ações de controle interno, corretivas e ouvidoria. Além disso, o Órgão é responsável por medidas para promover a transparência e prevenir a corrupção. O desempenho das atividades CGU muitas vezes se faz por meio de consolidação de informações e do cruzamento e mineração de dados. Tais atividades necessitam de uma visão holística e integrada dos seus ativos de dados, porém, a estrutura de gestão de dados existente é elementar e não tem se mostrado efetiva para atender às necessidades CGU. Em razão disso, o Órgão criou, em 2014, um grupo de trabalho para estudar a implementação do processo PO2–Arquitetura da Informação do Cobit 4.1. O grupo apresentou uma proposta de arquitetura da informação, não implementada até o momento, estruturada na forma de um modelo Entidade-Relacionamento (ER) que contemplava metadados das bases de dados do Órgão. No entanto, havia algumas lacunas nesse modelo, as quais motivaram a primeira etapa deste trabalho: uma revisão sistemática explorando o uso de ontologia na gestão de dados. Os resultados obtidos sugeriram a proposta de adoção de uma abordagem ontológica de gestão de dados e alicerçaram a proposta de modelo para arquitetura da informação subjacente, que se mostrou capaz de trazer melhorias ao modelo originalmente proposto. A principal contribuição deste trabalho para a academia e para outros Órgãos de governo é expor alguns benefícios de se adotar a abordagem ontológica na gestão de dados, com demonstração de aplicabilidade por meio de um protótipo para os domínios de dados da CGU.

Palavras-chave: Gestão de Dados, Ontologia, Integração, Arquitetura da Informação.

Abstract

This work studies the ontology as an approach to the data management and integration, it is considered as a mean to overcome the complexity involved in environments with variety of data sources. The varied applicability of the ontology directed the investigations of this work, which considered it as a base for the solution of the case study: CGU data management information architecture. The Ministry CGU is the arm of the federal executive branch with primary responsibility for internal control, corrective and ombudsman actions. In addition, the agency executes measures to promote transparency and prevent corruption. The performance of its various activities is often done through the consolidation of information, data cross-referencing and mining. Such activities require a holistic and integrated data assets view, nevertheless the existing data management structure is elementary and has not shown to be effective in meeting CGU requirements. As a result, in 2014 the Agency set up a working group to study the Cobit 4.1 PO2 - Information Architecture process implementation. The group presented an information architecture proposal, not implemented until now, structured in the form of a metadata Entity-Relationship (ER) model. However, there were some shortcomings in this model, which motivated the first stage of this work: a systematic review exploring the use of ontology in data management. The obtained results suggested the ontology based data management approach and supported the underlying information architecture proposed, which proved be able to bring about CGU model improvements. The main contribution of this work to the academy and other government offices is to present some ontology based data management benefits, with applicability demonstration by means of CGU data domains prototype.

Keywords: *Data Management, Ontology, Integration, Information Architecture.*

Sumário

1	Introdução	1
1.1	Problema de Pesquisa	1
1.2	Justificativa	3
1.3	Objetivos	4
1.3.1	Objetivo Geral	4
1.3.2	Objetivos Específicos	4
1.4	Resultados Esperados	5
1.5	Estrutura do trabalho	5
2	Fundamentação Teórica	6
2.1	Ontologia	6
2.1.1	Vocabulários e Ontologias	6
2.1.2	Tipos de ontologias	7
2.1.3	Desenvolvimento de Ontologias	7
2.1.4	Modelo Conceitual Suportado por Ontologia	8
2.1.5	Ontologia e Ciência da Computação	9
2.2	A Integração e Compartilhamento de Dados e Informações	9
2.2.1	Dados e Informações	9
2.2.2	Integração e Compartilhamento de Dados	9
2.2.3	Dados Mestres	10
2.3	Abordagem Ontológica de Gestão de Dados - OBDM	11
2.4	Cobit 4.1 e 5.0	11
2.5	<i>The Open Group Architecture Framework</i> - Togaf	12
2.6	Síntese do Capítulo	15
3	Mapeamento Sistemático	16
3.1	Questões de Pesquisa	16
3.2	O Processo de Mapeamento	17
3.3	Identificação dos Estudos Primários	17

3.4	Seleção	19
3.4.1	Crítérios de Inclusão e Exclusão	19
3.5	Coleta de Dados	21
3.6	Análise de Dados	22
3.6.1	Análise Inicial	22
3.6.2	Resultados e Análises	23
3.6.3	Respostas às Questões de Pesquisa	24
3.7	Discussão dos Resultados	27
3.8	Trabalhos Relacionados	28
3.8.1	Abordagem Ontológica de Gestão de Dados	28
3.8.2	Mapeamento entre Fontes e Ontologia	29
3.8.3	Integração de Dados Baseada em Ontologia	30
3.9	Síntese do Capítulo	32
4	Desenvolvimento das Arquiteturas e Estudo de Caso	33
4.1	Metodologias de Desenvolvimento de Arquiteturas	33
4.1.1	Abordagem <i>Enterprise</i> de Desenvolvimento de Ontologia	33
4.1.2	Método ADM - Fase A - para a Arquitetura de Dicionário de Dados	35
4.2	Contexto do Estudo de Caso	37
4.2.1	O Meta-Modelo PO2 da CGU	37
4.2.2	Iniciativas no Governo Federal	39
4.3	Modelo de Arquitetura da Informação	47
4.4	Arquitetura de Solução de Dicionário de Dados	53
5	Validação e Análise	56
5.1	Mapeamento entre Fontes e Classes	56
5.1.1	Resultados dos Mapeamentos	57
5.2	Avaliação do Modelo Conceitual de Dados	58
5.2.1	Perguntas do Questionário	60
5.2.2	Resultados da Avaliação	61
5.3	Análise	66
5.4	Síntese do Capítulo	67
6	Conclusão e Trabalhos Futuros	68
6.1	Contribuições	69
6.2	Trabalhos Futuros	69
	Referências	71

Lista de Figuras

2.1	Ciclo de Desenvolvimento de Arquitetura - ADM [1]	13
2.2	Visão Geral do Modelo de Conteúdo [1]	14
3.1	Processo de Revisão	17
3.2	Distribuição dos Artigos por Ano da Publicação	22
3.3	Distribuição dos Estudos por Categorias	25
3.4	Estrutura Ontológica da Arquitetura de Referência [2]	31
4.1	Meta-modelo de Arquitetura da Informação [3]	38
4.2	FACIN - Modelo Conceitual de Referência de Dados [4]	42
4.3	Representação dos Termos do VCGE em Forma de Árvore	44
4.4	Estrutura do modelo baseado em ontologia	48
4.5	Classes da Camada de Integração MD	49
4.6	Classes da Camada de Domínios de Negócio DN	50
4.7	Classes da Camada de Tarefas ou Serviços Gerais TG	51
4.8	Classes da Camada de Aplicações AP	52
4.9	Arquitetura de Solução de DD	53
5.1	Distribuição das Manifestações por Pergunta da Primeira Seção	62
5.2	Resumo das Manifestações às Questões da Primeira Seção	64
5.3	Distribuição das Manifestações por Pergunta da Segunda Seção	65
5.4	Distribuição das Manifestações às Questões da Segunda Seção	66

Lista de Tabelas

3.1 Resultados Preliminares	18
3.2 Resultado da Busca Refinada	19
3.3 Estudos Seleccionados	20
3.4 Distribuição dos Trabalhos Seleccionados por Publicação	23
3.5 Distribuição das Soluções Propostas nos Trabalhos	25
4.1 Propósito do Modelo Baseado em Ontologia	35
4.2 Elementos da Arquitetura de Solução de DD	54
5.1 Resumo das Fontes de Dados Mapeadas	57
5.2 Participantes da Avaliação	59
5.3 Questionário - Perguntas da Primeira Seção	60
5.4 Questionário - Perguntas da Segunda Seção	61

Capítulo 1

Introdução

O conceito de ontologia tem sido utilizado em vários trabalhos relacionando-o à organização e conhecimento de dados. Este conceito tem recebido especial atenção nos últimos anos, particularmente como abordagem para a arquitetura e gestão da informação, para a web semântica e para a mineração e integração de dados. Existem diversos artigos que indicam essa abordagem como meio de dominar a complexidade envolvida em ambientes com grande diversidade de fontes de dados. Os trabalhos apresentados por Daraio [5], Li [6] e Fitzpatrick [2], por exemplo, apresentam benefícios gerados pela compreensão da semântica dos dados alcançadas pelo uso de ontologias.

A aplicabilidade variada da ontologia direcionou as investigações deste trabalho no sentido de considerá-la como base da solução do estudo de caso apresentado. As questões de pesquisa que conduzem este trabalho tiveram como tema central a verificação do estado da arte no emprego da ontologia na gestão e integração de dados. Houve uma etapa inicial na qual foram analisados diferentes aspectos de pesquisas que se utilizaram da ontologia para compreender e organizar dados e informações, cujos resultados foram utilizados como base para desenvolver a proposta apresentada neste trabalho.

1.1 Problema de Pesquisa

O gerenciamento eficaz e eficiente em ambientes de informação complexos é um desafio, muitas vezes envolve o entendimento de conceitos pertencentes a diversos domínios de interesse bem como a relação desses conceitos com os sistemas de informação subjacentes. Organizar uma estrutura capaz de auxiliar o domínio dessa complexidade e de induzir a integração dos ativos de dados é uma condição necessária para o sucesso das ações de gestão nesse tipo de ambiente.

A CGU - Ministério da Transparência e Controladoria-Geral da União - encaixa-se no cenário descrito. Ela está estruturada em quatro unidades finalísticas: Secretaria de

Transparência e Prevenção da Corrupção (STPC), Secretaria Federal de Controle Interno (SFC), Corregedoria-Geral da União (CRG) e Ouvidoria-Geral da União (OGU) [7]. A atuação do Órgão é diversificada, mas em qualquer uma das suas unidades o principal insumo para os trabalhos são dados e informações. Por vezes, há necessidade de consolidação de informações do Governo Federal para disponibilização em aplicações web aos cidadãos; de cruzamento de dados para, por exemplo, identificar ilícitos; de mineração de dados para indicar, por exemplo, pontos de atenção nas auditorias a serem executadas.

Há três principais grupos de fontes de dados: sistemas próprios da CGU que dão suporte às suas atividades finalísticas; bases de dados de sistemas estruturantes do Governo Federal; e dados externos provenientes de sistemas de instituições ligadas ao Governo Federal ou, até, a outras esferas. O fato de o Órgão trabalhar tanto com dados gerados por aplicações próprias como por sistemas do Governo Federal, utilizados por todos os órgãos, gera um ambiente rico em dados e informações, porém complexo. A CGU mantém mais de 91 aplicações, que utilizam uma grande quantidade e diversidade de fontes de dados. Somente em banco de dados há mais de 6.800 tabelas, o que exige uma estrutura de gestão robusta para que esses ativos possam ser utilizados de maneira eficaz e eficiente.

A realidade é que a estrutura existente no Órgão é elementar, não fornece metadados - genericamente definidos como dados a respeito dos dados - organizados de forma a permitir uma visão abrangente e integrada desses ativos. Por esse motivo há perguntas básicas que são de difícil resposta, mesmo pelo administrador dos dados, por exemplo:

- **Q1.** Quantas e quais são as fontes de dados que contêm informações sobre a estrutura organizacional de governo, independentemente da esfera ou de poder ?
- **Q2.** Quais as fontes de dados das aplicações próprias CGU que lidam com dados de cidadãos e que poderiam ser integradas ?
- **Q3.** Em quais aplicações CGU e respectivas estruturas de dados é necessário efetuar pesquisas para se obter informação de todas as ações de controle que foram efetuadas sobre uma determinada instituição supervisionada pela CGU ?

Ademais, as lacunas de gestão existentes redundam em uma proliferação de dados inapropriada. Mais especificamente, ocorrem os seguintes problemas:

- **PC1. Redundância desnecessária**, resultante de replicações desnecessárias de algumas fontes de dados;
- **PC2. Arquitetura em "Silos"** dos sistemas de informação. Daraio [8] explica que a arquitetura baseada em silos ocorre em organizações de médio e grande porte e é constituída por várias fontes de dados independentes e distribuídas, cada uma servindo a uma aplicação específica.

As lacunas de gestão de dados resultam em dificuldades para localizar, entender, controlar e integrar dados. Em última instância, os trabalhos das áreas finalística como a consolidação de informações e o cruzamento e a mineração de dados ficam prejudicados. Há consequências indesejáveis, também, para os trabalhos da área de tecnologia da informação (TI), como a baixa interoperabilidade entre sistemas de informação.

A par dessa situação, o Órgão criou, em 2014, um grupo de trabalho para propor a estruturação do processo PO2 - Definir a Arquitetura da Informação do Cobit 4.1 [9], na expectativa de que, com a sua implementação, o Órgão possa superar as suas dificuldades de gestão de dados.

Esse processo Cobit possui dois objetivos de controle mais diretamente relacionados ao entendimento e ao compartilhamento de dados [9]:

1. **PO2.1** - Modelo de Arquitetura da Informação da Organização - que deve estabelecer um modelo de informação da organização, permitindo o desenvolvimento de aplicações e atividades de apoio à decisão consistentes com os planos de TI;
2. **PO2.2** - Dicionário de Dados Corporativo e Regras de Sintaxe de Dados- que mantenha um dicionário de dados, permitindo: o compartilhamento dos elementos de dados entre aplicativos e sistemas; o entendimento comum de dados entre TI e negócio; prevenir a criação de elementos de dados incompatíveis.

O produto apresentado pelo grupo de 2014 foi um modelo chamado Meta-Modelo de Arquitetura da Informação, cujo conteúdo considerou primordialmente metadados sob o ponto de vista de banco de dados. Os metadados do banco de dados são dados sobre um banco de dados - BD, que descreve a estrutura lógica e outras informações relevantes sobre uma fonte de dados. É a informação do esquema do BD, dados sobre o SGBD e dados técnicos relevantes necessários para acessar a fonte de dados [10].

Com não houve abordagem de aspectos conceituais dos ativos de dados, os quais são essenciais para facilitar a criação, o uso e o compartilhamento de dados e informações, há oportunidade de melhorias no Meta-Modelo proposto. As melhorias incluem a inserção de elementos semânticos no modelo e, também, uma estrutura adequada de dicionário de dados na qual o modelo será implementado.

1.2 Justificativa

Do ponto de vista científico, há interesse no tema ontologia para a gestão e integração de dados e na aplicação de modelos baseados nessa abordagem. Foi realizada uma revisão sistemática de literatura, a qual identificou 271 trabalhos sobre o tema entre 2006

e 2017. A curva de tendência da distribuição desses trabalhos por ano se mostrou crescente. Observou-se um bom número de publicações com proposta de modelo baseado em ontologia acompanhado de indicação de trabalhos futuros para uso e evoluções desses modelos.

Do ponto de vista da CGU, a proposição de modelos arquiteturais sob uma abordagem ontológica de gestão de dados pode contribuir para o atendimento dos requisitos do processo PO2 do Cobit 4.1 de entendimento e compartilhamento de dados.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo geral deste trabalho é identificar e entender os modelos de arquitetura da informação cuja estrutura base de organização são ontologias e que podem ser empregados na gestão e integração de dados. Há interesse especial nos modelos do domínio governamental e cujo foco é o estabelecimento de um entendimento comum sobre os dados corporativos, a fim de que possam ser utilizados em estudo de caso da CGU. O objetivo secundário é propor uma abordagem ontológica de gestão de dados para a CGU, materializada em modelos arquiteturais que auxiliem o atendimento dos objetivos de controle PO2.1 e PO2.2 do Cobit 4.1 [9].

1.3.2 Objetivos Específicos

Para atingir o objetivo geral deste trabalho, foram definidos os seguintes objetivos específicos:

- Realizar uma revisão sistemática de literatura para identificar artigos científicos sobre o objeto de estudo;
- Identificar os modelos de gestão de dados baseados em ontologia com potencial de uso direto ou como referenciais no atendimento das necessidades da iniciativa PO2 CGU;
- Identificar e analisar as iniciativas da Administração Pública Federal relacionadas à gestão de dados;
- Desenvolver uma proposta de Arquitetura da Informação (AI), na forma de protótipo de modelo de dados conceitual em camadas;
- Mapear as tabelas de banco de dados e os elementos do modelo conceitual em camadas;

- Desenvolver uma proposta de arquitetura de solução para a implementação do Dicionário de Dados (DD) corporativo, que seja capaz de armazenar metadados organizados a partir da AI proposta;
- Validar as propostas de AI e DD por meio de análise comparativa com a proposta CGU de 2014.

1.4 Resultados Esperados

O resultado esperado com o desenvolvimento deste trabalho é um entendimento mais amplo sobre os aspectos das abordagens ontológicas utilizadas na organização de metadados e seus efeitos no entendimento e na integração de dados. Tal entendimento poderá servir de subsídio às decisões sobre a abordagem de gestão de dados e o respectivo modelo de arquitetura da informação a serem adotados na CGU. Além disso, as explicações sobre o desenvolvimento e a análise do protótipo do modelo de arquitetura da informação proposto para o Órgão visam demonstrar a aplicabilidade da proposta apresentada bem como os benefícios em termos de entendimento e integração de dados e informações.

1.5 Estrutura do trabalho

Este trabalho está organizado em 4 capítulos, além deste. O Capítulo 2, apresenta os conceitos de Ontologia e da Gestão de Dados Baseada em Ontologia. Para melhor endereçar as necessidades do estudo de caso apresentado, alguns trechos do Cobit 4.1 e 5.0 e do *Framework* Togaf são destacados .

O Capítulo 3 apresenta a revisão sistemática de literatura realizada.

O Capítulo 4 descreve de maneira geral as metodologias escolhidas para a construção dos modelos propostos neste trabalho e, em seguida, o contexto e a solução proposta para o estudo de caso CGU.

O Capítulo 5 apresenta as validações do modelo de AI proposto, que ocorreram por meio de mapeamento de fontes de dados e de aplicação de questionário de avaliação.

O Capítulo 6 apresenta as conclusões deste trabalho, bem como os trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Este Capítulo apresenta uma revisão dos principais conceitos relacionados à ontologia e à integração e compartilhamento de dados e informações. Como a natureza do problema tratado neste estudo é de gestão e governança de dados, são apresentados alguns tópicos do *framework* de governança de TI Cobit e do *framework* de arquitetura corporativa Togaf.

2.1 Ontologia

De acordo com Borst [11] uma ontologia é definida como uma **especificação formal e explícita** de uma **conceitualização compartilhada**, na qual a especificação formal quer dizer algo que é legível para os computadores; explícita são os conceitos, propriedades, relações, funções, restrições e axiomas explicitamente definidos; conceitualização representa um modelo abstrato de algum fenômeno do mundo real e compartilhada significa conhecimento consensual.

Uma conceitualização é uma visão de mundo, uma maneira de pensar sobre algum domínio, e pode ser vista como um conjunto de regras informais que restringem a estrutura de uma realidade. Geralmente é expressa por um conjunto de conceitos (exemplo: entidades, atributos e processos), suas definições e seus inter-relacionamentos [12].

2.1.1 Vocabulários e Ontologias

Kohler [10] faz uma distinção entre vocabulário controlado e ontologia, na qual considera o primeiro listas de termos que estão bem definidos e podem ter um identificador. Nessa concepção, os elementos de um vocabulário controlado são chamados de conceitos e podem ser definidos de forma implícita ou enumerada explicitamente. Os termos ou identificadores de vocabulários controlados são frequentemente usados como entradas de

banco de dados. Definição: vocabulário controlado CV: = conjunto de conceitos nomeado c , com c : = (termo, definição, identificador, sinônimos). Um exemplo de vocabulário controlado é a Nomenclatura Enzimática (NC-IUBMB, 1992) [10].

Uma ontologia, por outro lado, consiste em conceitos que estão interligados por bordas direcionadas, formando assim um gráfico. As bordas de uma ontologia especificam de que maneira, por exemplo, '*Is-a*' ou '*Part-Of*', os conceitos estão relacionados entre si [10]. O modelo conceitual de dados proposto neste trabalho utiliza principalmente relações '*Is-a*' entre conceitos formando uma hierarquia.

2.1.2 Tipos de ontologias

Uma das classificações de ontologia amplamente utilizada foi proposta por [13]; Essa proposta é composta de quatro categorias, as quais são baseadas no nível de generalidade:

- **Ontologias de Topo ou de Fundamentação** – São ontologias gerais. Elas descrevem conceitos gerais, como elementos da natureza, espaço, tempo, coisas, estados, eventos, processos ou ações independentemente de um domínio ou problema específico.
- **Ontologias de Domínio** – Elas descrevem conceitos e vocabulários relacionados a um domínio específico, como medicina ou computação. É o tipo mais comum de ontologia, usualmente desenvolvida para representar um "mini mundo"[14].
- **Ontologias de Tarefas** – Elas descrevem tarefas genéricas ou atividades que podem contribuir para a resolução de um problema independentemente do domínio na qual ocorram, como o processo de vendas ou diagnóstico [14].
- **Ontologias de Aplicação** – Elas 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 tarefa correspondentes. Esses conceitos normalmente correspondem a regras aplicadas a entidades de domínio enquanto executam determinada tarefa[14].

Essa categorização é referencial para o modelo de dados em camadas proposto neste trabalho e apresentado no Capítulo 4.

2.1.3 Desenvolvimento de Ontologias

De acordo com Jones [15] há um número crescente de metodologias que abordam o desenvolvimento e manutenção de ontologias. Uma das metodologias citadas pelo autor

[15] é a "*Enterprise*". A abordagem é um modelo de construção baseado em etapas, apropriado para quando há objetivos e requisitos claros para a ontologia. Ela é baseada em quatro fases: identificação do propósito, identificação do escopo, formalização e avaliação formal. A identificação de propósito permite a determinação do nível de formalidade da descrição da ontologia e a identificação do escopo produz uma especificação de acordo com o domínio que a ontologia precisa representar. A identificação do escopo pode ser feita, por exemplo, utilizando questões de competência ou produzindo uma lista de conceitos potencialmente relevantes. As fases de formalização e de avaliação formal preocupam-se em criar definições formais e avaliar a ontologia utilizando, por exemplo, questões de competência.

As duas primeiras fases (identificação de propósito e de escopo) compõem uma etapa informal da metodologia e está concentrada em identificar e definir os conceitos chave da ontologia, o que a tornou particularmente interessante para o desenvolvimento do protótipo desenvolvido neste trabalho. Além disso, esta metodologia é considerada pertinente para o propósito deste trabalho por ser baseada em larga experiência de desenvolvimento de "enterprise ontology"[16], por tratar de domínio organizacional. Mais especificamente, utilizaram-se "Questões de Competência" [17] como guia para a identificação do propósito e a avaliação do modelo conceitual de dados proposto. As Questões de Competência são perguntas que se pretende responder a partir de inferências feitas na ontologia.

2.1.4 Modelo Conceitual Suportado por Ontologia

Os modeladores são livres para conceitualizar domínios da realidade da maneira que eles gostem ou que melhor se ajuste aos seus propósitos. Essa liberdade traz seus problemas, já que diferentes usuários podem modelar o mesmo domínio de maneiras diferentes e de acordo com diferentes conceitualizações. Além disso, as linguagens usadas na especificação de modelos conceituais como *Unified Modeling Language* - UML, *Entity-Relationship Diagram* - ER e *Role Modeling* - ORM são frequentemente linguagens orientadas a software, ou seja, mais próximas da implementação em software e não expressivas o suficiente para capturar completamente o conhecimento necessário sobre o domínio. Como consequência, a falta de semântica nos modelos criados com essas linguagens leva a problemas como interpretações equivocadas, representações ambíguas e imprecisas e dificuldades na comunicação [18].

Para remediar essa situação, na última década surgiu o que é conhecida como modelagem conceitual suportada por ontologia, que consiste na prática de usar ontologias e princípios ontológicos para apoiar a atividade de modelagem conceitual [19], [20], [21], [22].

2.1.5 Ontologia e Ciência da Computação

A ontologia é usada para melhorar a comunicação humana ou entre computadores. De maneira concreta, a ontologia é usada para auxiliar a comunicação entre agentes humanos, para atingir a interoperabilidade entre sistemas computacionais, ou para aprimorar processos ou melhorar a qualidade de produtos de engenharia de software [23] [12]. Para [12], na interoperabilidade a ontologia é usada como um inter método de conversão, que pode fazer tradução entre diferentes paradigmas, linguagens e ferramentas de software.

Afora isso, há um benefício de interesse especial: o uso de ontologias para o estabelecimento de banco de dados semânticos, os quais possibilitam interfaces para usuários efetuarem pesquisas em bases de dados sem conhecimento da estrutura ou detalhe técnico a respeito das fontes de dados [10].

2.2 A Integração e Compartilhamento de Dados e Informações

2.2.1 Dados e Informações

Antes de se abordar o assunto integração e compartilhamento de dados e informações, é importante esclarecer que em muitos tópicos desta trabalho os termos 'dados' e 'informações' são utilizados sem distinções, particularmente nas situações em que a diferença entre eles não é significativa para o conteúdo abordado. Nos trechos do trabalho em que a distinção for necessária, ela será destacada e terá como base as seguintes conceituações de referência: 1. 'dado é um conjunto de registros qualitativos ou quantitativos, conhecido, que organizado, agrupado, categorizado e padronizado adequadamente transforma-se em informação'; 2. 'informação são dados organizados de modo significativo, sendo subsídio útil à tomada de decisão' [24].

2.2.2 Integração e Compartilhamento de Dados

Os dados organizacionais, no sentido estrito do conceito 1, geralmente estão localizados em sistemas especializados, desenvolvidos para atender necessidades específicas, formando silos. Segundo [25], esses silos são difíceis de unir para fornecer visualizações transversais dos dados. Existe uma necessidade crescente de fornecer dados entre domínios, uma tarefa geralmente muito difícil, considerando que raramente há qualquer convenção semântica comum que permita a interoperabilidade entre sistemas.

Uma das formas usuais de superar a barreira criada por sistemas em silos, e viabilizar o fornecimento de informação multidomínio, é a integração de dados. A definição de

integração de dados utilizada neste trabalho é a de Lenzerini [26], que, de forma holística, considera a integração de dados como a capacidade computacional de abordar o problema de fornecer dados localizados em diversas fontes heterogêneas sob uma perspectiva única. Neste trabalho, a integração é considerada a principal forma de compartilhamento de dados.

Segundo [27], há basicamente duas abordagens para se revolver o problema de integração de dados: a procedural e a declarativa. Na primeira abordagem os dados são integrados de maneira *ad-hoc* em relação a um conjunto de necessidades de informação predefinidas. Nesse caso, são projetados módulos que encapsulam os dados originais das fontes expondo-os em um formato de modelo comum predefinido. Em seguida, módulos mediadores recebem dados ou informações de um ou mais módulos encapsuladores e fazem a integração e a resolução de conflitos para prover informações refinadas para outros mediadores ou usuários finais.

Na segunda abordagem, declarativa, o objetivo é modelar os dados nas fontes utilizando uma linguagem adequada a fim de construir uma representação unificada dos dados a ser referenciada para pesquisas futuras.

Um grande diferencial entre as duas abordagens está no fato de a primeira não necessitar o conhecimento do esquema de dados das fontes, enquanto que a segunda utiliza os esquemas de dados para a modelagem da representação unificada. Mesmo que diferentes, ambas abordagens requerem o entendimento abrangente das entidades transversais e dos domínios de negócio da(s) organização(ões) envolvida(s), para estabelecer o formato de modelo predefinido e para a representação unificada dos dados, respectivamente. O modelo de arquitetura da informação a ser proposto para a gestão de dados CGU deverá dar suporte tanto à abordagem procedural quanto à declarativa.

2.2.3 Dados Mestres

Dados Mestres organizacionais ou corporativos referem-se a temas comuns de dados que são transversais a quaisquer negócios e aos objetos de informação subjacentes. Em qualquer organização, há conceitos comumente reconhecidos como foco dos processos de negócios. Por exemplo: clientes, produtos, fornecedores partes, locais, perfis, itens contábeis, contratos e políticas [28].

Dados Mestres são "coisas" importantes (objetos de negócio) que são registradas nos sistemas transacionais e tendem a existir em mais de uma área de negócio dentro das organizações. Sob um objeto dado mestre pode existir uma relação de hierarquia explícita ou implícita [28]. Em função do requisito de compartilhamento de dados, é fundamental que a solução a ser proposta pra o caso CGU dê especial atenção aos dados mestres do Órgão.

2.3 Abordagem Ontológica de Gestão de Dados - OBDM

O conceito de ontologia, de acordo com [8], é a noção central da abordagem *Ontology-Based Data Management Approach* – OBDM - e a reflexão sobre ela é a base de todas as atividades de um sistema OBDM. Os autores esclarecem que as ideias sobre essa abordagem foram introduzidas em Calvanese [29], Poggi [30] e Lenzerini [31]. Elas surgiram de diversas disciplinas, em particular, da Integração da Informação, da Representação da Lógica e do Conhecimento e de Banco de Dados Dedutivos. O trabalho [32] apresenta a abordagem ontológica para a Gestão de Dados (OBDM) como meio de alcançar a integração de dados e a interoperabilidade entre sistemas.

O objetivo da abordagem OBDM de gestão de dados voltada à integração é totalmente aderente ao objeto de estudo deste trabalho, assim, as ideias dessa abordagem serão direcionadoras das análises e propostas apresentadas.

2.4 Cobit 4.1 e 5.0

Os objetivos de controle **PO2.1 – Modelo de Arquitetura da Informação da Organização** e **PO2.2 – Dicionário de Dados Corporativos e Regras de Sintaxe de Dados** do Cobit 4.1[9], que tratam das saídas foco deste trabalho, são diretamente mapeados para a prática de gestão **APO03.02 Definir a arquitetura de referência** do processo **AP003 – Gerenciar a Arquitetura Corporativa** do Cobit 5.0 [33].

A prática de gestão APO03.02 possui atividade específica destinada à manutenção do modelo de Arquitetura da Informação (AI) e à manutenção do dicionário de dados corporativo (DD). O modelo de AI e o DD devem permitir, entre outros, o uso otimizado da informação para a tomada de decisões e o compartilhamento otimizado da informação entre aplicações, por meio de um entendimento comum dos dados. Há praticamente uma reescrita, com adições, dos objetivos de controle detalhados nos processos **PO2.1 e PO2.2** do Cobit 4.1 [9].

Para tais atividades, de acordo com o descrito no processo **APO03 – Orientação Relacionada** [33], o processo **AP003 – Gerenciar a Arquitetura Corporativa** utiliza o *Framework* Togaf 9 como referência [1].

Os processos Cobit são utilizados como referência de governança de TI em várias organizações, inclusive na CGU. Os investimentos do Órgão para atender aos objetivos de controle do processo PO2, ainda que da versão 4.1 do *framework*, são válidos, pois eles seriam aproveitados como práticas de gestão em caso de evolução para versão mais atual do *framework*.

2.5 *The Open Group Architecture Framework - Togaf*

Segundo Jonkers e Lankhorst [34], uma definição comumente utilizada de arquitetura no escopo de Tecnologia da Informação - TI é a do IEEE Standard 1471-2000 [35]: a arquitetura é a organização fundamental de um sistema materializada em seus componentes, as relações entre seus componentes e entre componentes e o meio ambiente. Essa materialização inclui, também, os princípios orientadores de seu *design* e evolução.

O Togaf é um modelo conceitual de arquitetura corporativa concebido em 1995 pelo *The Open Group Architecture Forum*, cujo objetivo é fornecer uma abordagem global para o desenho, o planejamento, a implementação e a governança de arquiteturas. Atualmente, na versão 9.1, publicada em 2011, o Togaf se baseia em um processo iterativo, reutilizável, cíclico e suportado pelas melhores práticas de modelagem envolvidas nas atividades fim ou meio de uma organização, compreendendo quatro domínios: negócios, dados, aplicações e tecnologia [1].

Fundamentalmente, um *framework* de arquitetura é um conjunto de estruturas que pode ser usada para desenvolver uma ampla gama de arquiteturas diferentes. O *framework* de arquitetura corporativa deve descrever um método para projetar um estado futuro (alvo) de uma organização em termos de um conjunto de blocos de construção e deve mostrar como os blocos de construção se encaixam [1]. O Togaf provê o método de desenvolvimento de arquitetura chamado ADM – *Architecture Development Method*.

Método ADM de Desenvolvimento de Arquitetura

O ADM é um ciclo composto de fases conforme apresentado na Figura 2.1. As fases vão desde a fase "A" de visão inicial da arquitetura até a fase "G" de implementação da arquitetura proposta (alvo). Há uma fase posterior, "H", que identifica novas mudanças necessárias.

Cada fase do ciclo ADM é descrita por meio de seu objetivo, entradas, passos, saídas e abordagem. O Togaf frisa que o método deve ser customizado para cada organização ou até mesmo para atender a situações específicas. Nem tudo que está descrito é necessário para atingir o objetivo da arquitetura produzida e, de mesma forma, pode haver elemento adicional aos descritos que seja necessário numa situação específica.

Por exemplo, a fase "A" tem como objetivos, entre outros, criar uma visão da proposta de arquitetura que possa auxiliar na decisão de sua aprovação e, também, auxiliar o entendimento dos seus impactos. As saídas da fase "A" contemplam, entre outras, a Visão de Arquitetura e o Esboço do Documento de Definição de Arquitetura. A Visão de Arquitetura inclui, mas não está restrita a: Descrição do Problema; Visualizações Resumidas;

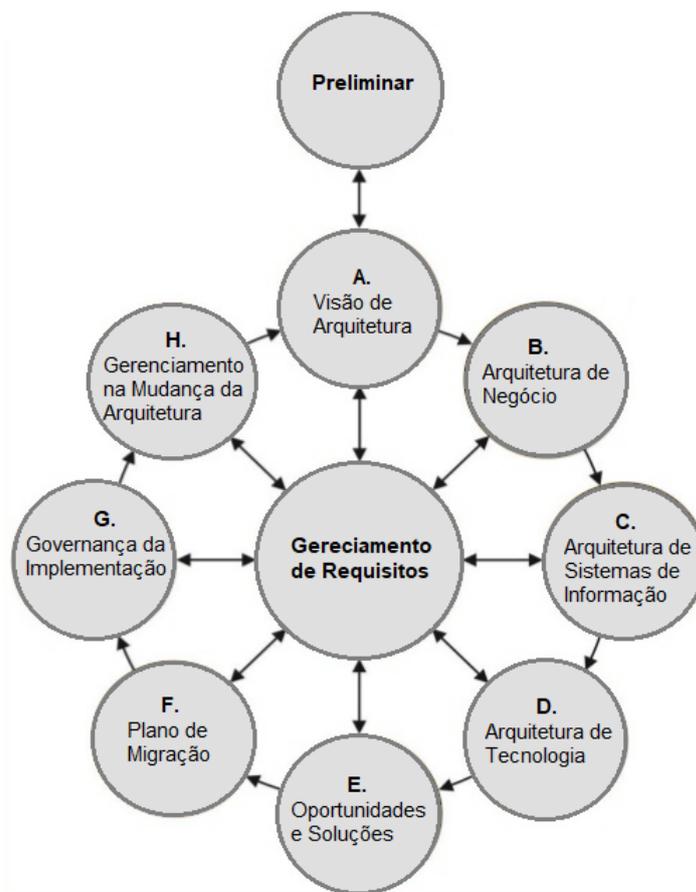


Figura 2.1: Ciclo de Desenvolvimento de Arquitetura - ADM [1]

Requisitos Principais de Alto Nível das Partes Interessadas. O Esboço do documento de definição de arquitetura, inclui, mas não está restrito a: Arquitetura(s) versão 0.1 de Linha de Base e Arquitetura(s) Alvo, contemplando um ou mais domínios Togaf.

Linguagem Archimate

Um dos padrões do Open Group Togaf é a especificação da ArchiMate, uma linguagem de modelagem de arquitetura corporativa cuja especificação livre e independente é suportada por diferentes fornecedores de ferramentas. A especificação da linguagem dá instrumentos para a descrição, análise e visualização de relacionamentos entre domínios de negócios. Ela é uma linguagem comum para a descrever a construção e operação de processos de negócio, estruturas organizacionais, fluxos de informação, sistemas de TI e infraestrutura tecnológica [36].

Modelo de Conteúdo Togaf

Uma das estruturas disponibilizada pelo *framework* Togaf é o seu Modelo de Conteúdo, organizado por domínios, conforme apresentado na Figura 2.2. Esse Modelo oferece uma visão abrangente dos elementos usualmente encontrados nas organizações, auxiliando a identificação dos blocos de construção que poderão fazer parte de qualquer arquitetura a ser desenvolvida. Ele é praticamente um conjunto de descrições relevantes para representar organizações de qualquer natureza.

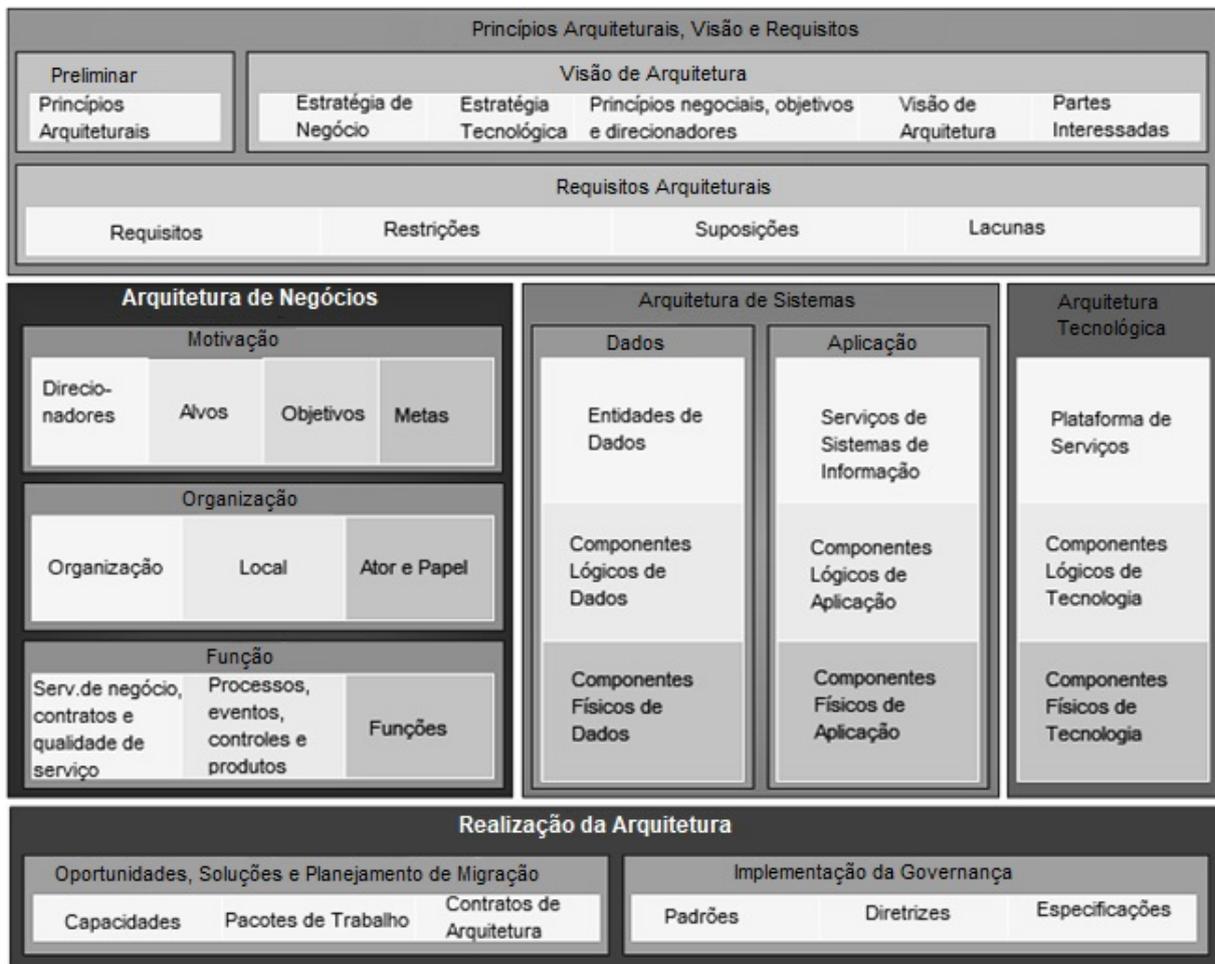


Figura 2.2: Visão Geral do Modelo de Conteúdo [1]

O processo ADM e o modelo de conteúdo Togaf serão considerados para o desenvolvimento das arquiteturas técnicas a serem apresentadas neste trabalho.

2.6 Síntese do Capítulo

Este Capítulo apresentou os conceitos básicos necessários para contextualizar as ideias utilizadas neste trabalho, bem como os tópicos complementares Cobit e Togaf, que compõem o contexto do estudo de caso apresentado. Além disso, houve uma breve explicação de como cada conceito ou tópico está relacionado ao desenvolvimento dos produtos apresentados.

Capítulo 3

Mapeamento Sistemático

Um mapeamento sistemático consiste em uma pesquisa da literatura para determinar quais estudos abordam uma questão definida, onde foram publicados, em que bibliotecas foram indexados, que tipo de resultados estes estudos retornaram [37], [38].

Preliminarmente, um tópico de interesse necessita ser definido para o processo de revisão. Em seguida, as questões de pesquisa relacionadas ao tópico selecionado são formuladas, pois elas compõem a base do estudo e servem de guia durante a coleta de dados.

As questões de pesquisa definidas na etapa de mapeamento sistemático deste trabalho são relacionadas à proposta OBDM e a outras semelhantes que adotem a ontologia como direcionador principal para o gerenciamento ou integração de dados. O objetivo que direcionou essa etapa, já concluída, foi de responder às questões definidas e identificar ideias que poderiam servir de referencial para aprimorar o modelo **PO2 CGU**.

3.1 Questões de Pesquisa

A fim de identificar evidências de que a utilização de abordagem ontológica na gestão de dados de governo poderia trazer bons resultados em termos de integração e entendimento dos dados, foram elaboradas quatro questões de pesquisa (**QP**):

QP1) Quais as categorias dos estudos que propõem o uso de ontologias para a gestão ou integração de dados, particularmente as relacionadas com assuntos de interesse do setor governo ?

QP2) Quais são os tipos de ontologias e estratégias de construção que são utilizados nas arquiteturas de informação ?

QP3) De acordo com os autores, quais tópicos nas abordagens tipo OBDM necessitam de pesquisa futura ?

QP4) Quais as limitações ou trabalhos em aberto da abordagem OBDM?

3.2 O Processo de Mapeamento

O processo de revisão/mapeamento sistemático de literatura foi dividido em 4 passos ou etapas, de acordo com o apresentado na Figura 3.1. O primeiro passo preocupa-se em identificar as conferências e jornais que são candidatos a representar o estado da arte em pesquisa no tópico de interesse. Os estudos primários para revisão são pré-selecionados a partir das palavras-chave ainda nesse passo. O segundo passo visa a exclusão ou a inclusão dos artigos pré-selecionados com base na leitura do título e resumo de cada um. Depois, no terceiro passo, os textos selecionados são classificados e as informações acerca das questões de pesquisa são extraídas. O quarta e último passo efetua a análise dos dados coletados. O processo de revisão descrito é baseado no guia para revisões sistemáticas de literatura elaborado por Kitchenham e Charters [39].



Figura 3.1: Processo de Revisão

O objetivo de identificar algumas ideias para aprimorar ou propor novo modelo para a CGU gerou uma atividade extra no processo de mapeamento. Alguns textos que apresentaram modelos promissores para as necessidades CGU impulsionaram pesquisas complementares sobre as propostas apresentadas nesses textos. Assim, praticamente houve uma execução complementar, antes da execução do terceiro passo, dos passos iniciais do processo utilizando palavras-chave relacionadas aos modelos considerados promissores.

As próximas seções descrevem os resultados obtidos em cada etapa do processo.

3.3 Identificação dos Estudos Primários

O primeiro passo foi a identificação dos trabalhos que mencionam ideias que se assentaram em abordagens tipo OBDM com a seleção de fontes (bases de dados) apropriadas. As bibliotecas digitais:

- IEEE xplore – ieeexplore.ieee.org/;

- ACM – <http://www.acm.org/>;
- Springer – <https://link.springer.com/>;
- ScienceDirect – www.sciencedirect.com/;
- Computer and Information Systems - Proquest – <https://search.proquest.com/>.

Para garantir uma identificação abrangente e imparcial, determinadas palavras-chave foram aplicadas a fim de realizar o processo de busca. Essas palavras-chave foram utilizadas em pesquisas de teste a fim de verificar os retornos que poderiam ser obtidos.

Como resultado da estratégia de busca por testes, efetuada em *search engine* das bibliotecas selecionadas, foram obtidos os conjuntos de *string* de busca, conforme apresentado na Tabela 3.1, com resumo dos números preliminares obtidos:

Tabela 3.1: Resultados Preliminares

Termos da Busca	Springer	ACM	IEEE	Proquest	Science Direct
Data governance integration ontology	141	39	10	40	13
Data management integration ontology	657	331	482	425	698
Information management integration ontology	1658	299	538	436	544

Após isso, como refinamento, foram acrescentadas as *strings* "*Government*" e "*architecture*" a fim de restringir as pesquisas a esses domínios e também desprezado o conjunto de *strings* de governança de dados, pois foi considerado que o assunto governança é implementado com a gestão de dados.

Além disso, foram adicionados filtros presentes nas páginas das bibliotecas digitais para que se selecionasse somente as pesquisas que contivessem as *strings* "*data management*" ou "*information management*" em seu título ou sumário. A pesquisa de texto completo traria muitos resultados irrelevantes [39], [40]. O refinamento foi complementado com os seguintes critérios:

- Estudos publicados no período compreendido entre 2006 e 2017;
- Estudos que estão disponíveis nos periódicos (bases) da CAPES em inglês;
- A exclusão dos estudos classificados como somente de outras áreas do conhecimento (medicina, matemática, etc.).

O resultado obtido após a busca refinada é apresentado na Tabela 3.2. Observou-se que há 17 estudos que se encontravam em mais de uma biblioteca de busca. Retiradas as repetições, o total geral encontrado foi de 253 estudos.

Tabela 3.2: Resultado da Busca Refinada

Termos da Busca	Springer	ACM	IEEE	Proquest	Science Direct	Total
Data management integration ontology government achitecture	7	29	39	14	18	107
Information management integration ontology government architecture	3	44	70	26	21	164
Total	10	73	109	40	39	271

Os estudos foram classificados de acordo com os campos enumerados na Seção 3.4 para se construir um panorama sobre o tema.

1. Ano da publicação;
2. Jornal ou Conferência.

3.4 Seleção

A etapa de busca descrita na Seção 3.3 resultou em uma quantidade considerável de trabalhos/estudos, demandando a inclusão de novos critérios. Assim, foram analisados os títulos, as palavras-chave e os resumos. Uma leitura exploratória auxiliou na redução do número total de trabalhos selecionados, permanecendo somente os que apresentam indicativos de que poderiam ser relevantes para responder às questões de pesquisa. O principal critério de seleção foi a abordagem do tema ontologia na gestão de dados, principalmente com objetivo de integração. Qualquer das formas de tratamento do assunto (investigação de métodos, ferramental, teorias, exemplos, implementações, avaliações e etc.) foram consideradas interessantes. Ademais, a seleção foi restrita a artigos em razão do foco se dar em conhecer o andamento das pesquisas.

Após essa fase, foram acrescentados alguns artigos que demonstravam potencial para complementar a visão dos modelos de interesse para a CGU, independentemente de figurarem nos resultados preliminares das buscas. O novo universo de estudos (resultado das buscas seletivas mais estudos acrescentados) foi submetido a critérios de inclusão e exclusão, os quais são definidos na Seção 3.4.1.

3.4.1 Critérios de Inclusão e Exclusão

Os critérios de inclusão adotados (CI) foram:

CI.1) Estudos que contêm proposições de ontologias para gestão ou integração de dados.

CI.2) Estudos que propõem uma abordagem, processo ou metodologia para estabelecer o uso de ontologias na gestão e/ou integração de dados.

CI.3) Estudos com proposta de arquitetura para solução de repositório de metadados que considere o uso de ontologia.

CI.4) Estudos empíricos ou experimentais que utilizam o mapeamento semântico entre fontes de dados e ontologia como forma de entender e localizar informações para obter interoperabilidade entres sistemas.

Os critérios de exclusão (**CE**) adotados foram:

CE.1) Estudos fora do domínio do estudo primário.

CE.2) Teses (por exemplo, teses de mestrado) ou monografias.

CE.3) Estudos incompletos, como apenas resumos e resumos expandidos, que não permitem um entendimento de sua proposta.

CE.4) Estudos com menos de cinco páginas, publicados como *ShortPaper*.

Após a aplicação dos critérios mencionados, foram obtidos os 28 estudos listados na tabela 3.3 , número bem menor que o resultante da busca seletiva. A triagem mais rigorosa de estudos, com a intenção de identificar os estudos com relação mais próxima ao objetivo do **PO2 CGU**, não é limitante aos resultados esperados deste trabalho.

Tabela 3.3: Estudos Seleccionados

ID	Título	Ano
E1	A Holistic Approach for the Architecture and Design of an Ontology-Based Data Integration Capability in Product Master Data Management [2].	2012
E2	A Learner Oriented Ontology to Make Effective Learning Management Systems [41]	2006
E3	A new approach to a local e-government portal for information management and deep searching [42]	2006
E4	Achieving Interoperability Among Healthcare Standards: Building Semantic Mappings at Models Level [43]	2012
E5	An ontology-based enterprise architecture [44]	2010
E6	An ontology-centric architecture for extensible scientific data management systems [6]	2013
E7	An Ontology-oriented Decision Support System for Emergency Management Based on Information Fusion [45]	2015
E8	Automatic generation of ontology from data models: A practical evaluation of existing approaches [46]	2013

E9	Building an operational product ontology system [47]	2006
E10	Data integration for research and innovation policy: an Ontology-Based Data Management approach [32]	2016
E11	Extensive overview of an ontology-based architecture for accessing multi-format information for disaster management [48]	2012
E12	Metadata management, interoperability and Linked Data publishing support for Natural History Museums [49]	2014
E13	NGS: a framework for multi-domain query answering [50]	2008
E14	Ontology for Semantic Representation of Marine Metadata [51]	2013
E15	Ontology Techniques for Representing the Problem of Discourse: Design of Solution Application Perspective [52]	2016
E16	Ontology-Based Approach for Academic Evaluation System [53]	2017
E17	Ontology-based Data Management [31]	2011
E18	Ontology-Based Information Integration and Decision Making in Prefabricated Construction Component Supply Chain [54]	2017
E19	Ontology-driven data integration for railway asset monitoring applications [55]	2014
E20	OPSDS: A Semantic Data Integration and Service System Based on Domain Ontology [56]	2016
E21	OVM: An Ontology for Vulnerability Management [57]	2009
E22	Provenance ontology model for land administration spatial data supply chains [58]	2015
E23	Sapientia: the Ontology of Multi-dimensional Research Assessment [8]	2015
E24	Semantic enhanced cloud environment for surveillance data management using video structural description [59]	2016
E25	Semantic web based distributed government data center [60]	2011
E26	SEMEDA: ontology based semantic integration of biological databases [10]	2003
E27	Service and Model-driven Dynamic Integration of Health Data [61]	2011
E28	The advantages of an Ontology-Based Data Management approach: openness, interoperability and data quality [5]	2016

3.5 Coleta de Dados

A coleta dos dados foi baseada na leitura integral dos artigos selecionados. Os dados relevantes foram mapeados diretamente a partir das questões de pesquisa e foram registrados em uma tabela. Depois, as classificações dos trabalhos com relação às questões de pesquisa foram adicionadas [39].

3.6 Análise de Dados

3.6.1 Análise Inicial

A construção de um panorama da evolução e atenção recebida pelo tema nas pesquisas acadêmicas desde 2006 foi efetuada a partir dos resultados obtidos na busca refinada, tabela 3.2.

A distribuição dos trabalhos da busca refinada por ano de publicação encontra-se na Figura 3.2. As barras representam o número absoluto de textos identificados por ano. Em geral, observou-se atividades de pesquisa na área de ontologia para a gestão e integração de dados desde 2006 até o período do mapeamento sistemático. O pico de atividade observado ocorreu no ano de 2016 e o menor número encontrado foi no ano de 2007. A linha pontilhada indica a tendência no período, apontando para um crescimento nas pesquisas, mesmo que suave.

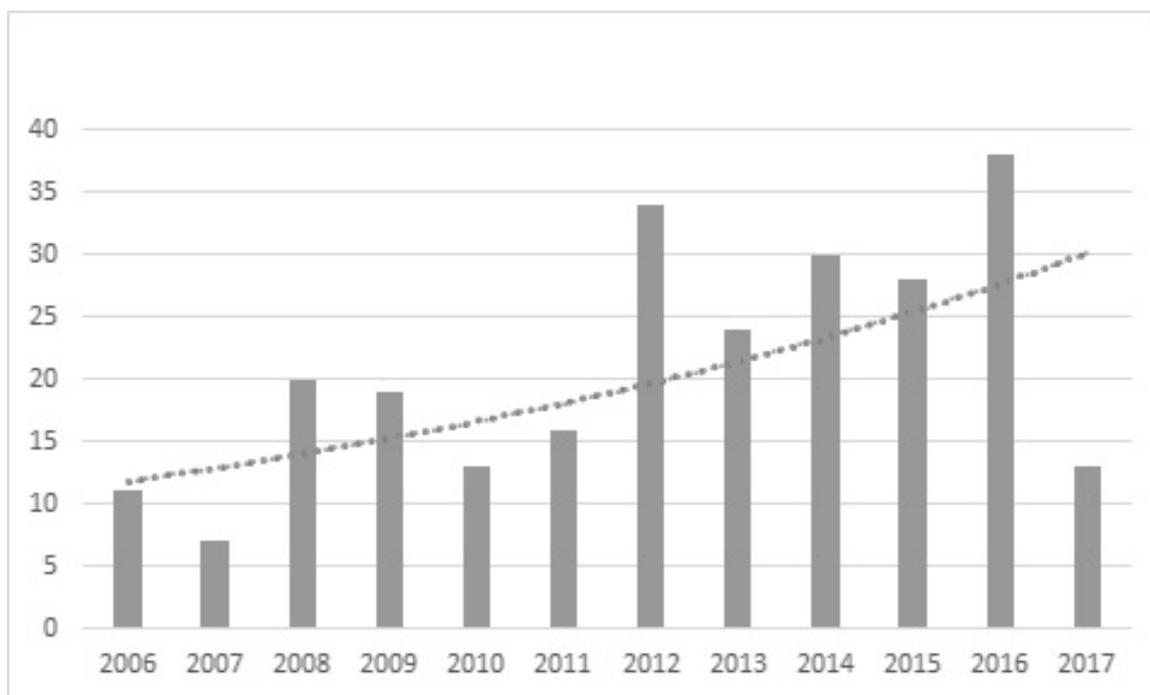


Figura 3.2: Distribuição dos Artigos por Ano da Publicação

A distribuição dos trabalhos selecionados na busca refinada por publicação encontra-se na tabela 3.4. Os dados de distribuição contemplaram somente as publicações com três ou mais artigos.

Observa-se que dentre as quatro publicações com maior número de artigos (com sete ou mais publicações) há três com tema *Information ou Data Management*. A única exceção nesse grupo tem tema sobre *Big Data*, que certamente remete aos estudos sobre

gestão de dados em função do tratamento necessário nesses conjuntos de dados muito grandes ou complexos (*Big Data*).

Tabela 3.4: Distribuição dos Trabalhos Seleccionados por Publicação

Jornal ou Conferência	Quant.
ACM Trans. Manage. Inf. Syst.	6
Advanced Engineering Informatics	4
Computers in Industry	6
Environmental Modelling & Software	4
IEEE International Conference on Data Engineering	5
IEEE International Conference on Big Data (Big Data)	11
IEEE International Conference on Information Reuse and Integration	3
IEEE International Congress on Big Data	3
Information Systems	4
Information Systems Frontiers	3
International Conference on Advanced Information Networking and Applications (aina)	3
International Conference on Data Engineering (ICDE'06)	3
International Conference on Digital Information Management	5
International Conference on Information Fusion	3
International Conference on Information Management, Innovation Management and Industrial Engineering	7
Journal of Information Systems and Technology Management : JISTEM	3
Proceedings of the ACM International Conference on Information and Knowledge Management	5
Proceedings of the ACM International Conference on Information and Knowledge Management	3
Proceedings of the ACM SIGMOD International Conference on Management of Data	7
Proceedings of the International Conference on Ubiquitous Information Management and Communication	10
Proceedings of the Workshop on Semantic Web Information Management	3

3.6.2 Resultados e Análises

A análise de dados resume os resultados da coleta de dados e fornece respostas às questões de pesquisa. A análise pode ser descritiva ou quantitativa, tanto uma quanto a outra devem ser organizadas em tabelas ou figuras relacionadas às questões de pesquisa. Tais tabelas ou figuras devem destacar as diferenças e as similaridades entre os textos. A respeito da classificação dos conteúdos dos textos considerou-se haver uma interação entre a coleta de dados e a sua análise. Por exemplo, a classificação das abordagens e ontologias

em tipos (QP2) foi baseada nas informações encontradas no decorrer da análise do texto [39], [40].

Há duas principais advertências com relação à validade de algumas conclusões. Primeiro, pode haver viés na seleção, a amostra escolhida pode não representar a situação atual das pesquisas. Mesmo assim, a seleção efetuada foi considerada representativa, a tabela 3.3, que lista os estudos selecionados, mostra uma base suficiente e diversificada de textos identificados abordando o estudo deste trabalho. A segunda ameaça recai sobre o processo de coleta de dados. Por exemplo, a classificação dos textos foi baseada, de certo modo, em decisão subjetiva. Tal fato foi amenizado pelo uso de questões de pesquisa, que auxiliaram no alcance de um nível adequado de objetividade.

As questões de pesquisa formuladas na Seção 3.1 foram utilizadas para estruturar os resultados encontrados.

3.6.3 Respostas às Questões de Pesquisa

Os estudos selecionados, apresentados na Tabela 3.3, foram a base para as análises que se seguem, referentes às questões QP1 a QP4.

QP1) Quais as categorias dos estudos que propõem o uso de ontologias para a gestão ou integração de dados?

Esta questão é relacionada aos objetivos principais dos estudos selecionados. A maioria dos estudos, 23 do total ou 82%, têm como um dos seus focos a proposição de modelos, quer empíricos ou analíticos.

As outras categorias de estudos que ocorreram, mas em um número bem reduzido, foram: procedimento/técnica e ferramenta. Os estudos E23 [8] e E8 [46] apresentam técnica de avaliação da qualidade de ontologias. No caso particular de E8 o objetivo é avaliar modelos gerados de forma semi-automática.

O estudo E26 [10] é o único que tem como um dos seus pontos centrais uma proposta de ferramenta, que inclui a edição de ontologias e vocabulários e o acesso semântico a dados. Sob esse aspecto, é interessante destacar que todos os outros trabalhos que mencionaram o uso de ferramenta de edição de ontologias, 7 no total, fizeram uso da ferramenta Protégé 4.3 [62] ou versão superior. Entre eles estão os estudos E3 [42], E11 [48], E15 [52], E16 [53], E18 [54], E22 [58] e E24 [59].

A distribuição por categorias é apresentada na Figura 3.3. Há estudos que têm diversos focos, o que implica a sua inclusão em mais de uma categoria.

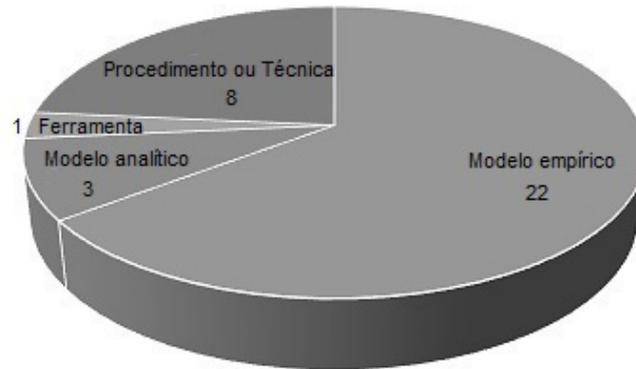


Figura 3.3: Distribuição dos Estudos por Categorias

QP2) Quais são os tipos de ontologias e estratégias de construção que são utilizados nas arquiteturas de informação?

Antes de responder à questão **QP1** foi elaborado um resumo dos tipos de arquiteturas propostas nos estudos, segmentando os estudos pelo objetivo principal para que se destinam. Foram encontrados 05 estudos cujo objetivo era de argumentar sobre as vantagens ou propor um método de desenvolvimento de soluções com uso de ontologia, por isso, não estão incluídos no total apresentado na Tabela 3.5. A coluna "solução técnica" indica o número de arquiteturas de solução técnica apresentadas nos estudos nas quais as ontologias serão utilizadas como base para a arquitetura da informação.

Tabela 3.5: Distribuição das Soluções Propostas nos Trabalhos

Objetivo principal do estudo	Arquiteturas Propostas		
	Ontologia-AI	Solução Téc.-ST	AI e ST
Abordagem ontológica			1
Metadados e gestão	1	1	4
Integração e Interoperabilidade	4	2	10
TOTAL = 23	5	3	15

Foram identificadas 20 propostas de ontologias apresentadas nos trabalhos, 15 das quais acompanhadas de proposta de solução técnica. Na maioria delas é apresentada uma proposta de ontologia de domínio específico. Porém, há menção a outros tipos de ontologia como: ontologia *global/local* nos estudos E11 [48], E16 [53] e E20 [56]; ontologias *Upper Domain, Top-Level* ou *Foundation* nos estudos E1 [2], E3 [42], E18 [54] e E19 [55]. Além disso, estes últimos estudos mencionam as ontologias do tipo *application, standard, multi-*

domain (data-integration), domain-independent concepts, task, light e content (portais). É interessante destacar que 17 modelos, 85%, foram formalizados em OWL [63].

Quanto à estratégia de desenvolvimento de ontologias para a arquitetura da informação, observou-se que a maioria dos estudos trabalharam em uma proposta de ontologia específica do domínio. Porém, nota-se que os estudos como E1 [2], E6 [6], E18 [54] e E19 [55], que trabalharam com níveis mais elevados de ontologias, como as de topo ou de multidomínio, sugerem uma estratégia *top-down* de arquitetura. Ainda houve casos de estudos em que foram mencionadas ontologias padrão de domínio (por exemplo o estudo E4 [43], E26 [10] e E27 [61]) e a estratégia utilizada foi a de estender tais modelos, podendo ser a partir de um único ou diversos modelos. Neste último caso, pode haver necessidade de *matching* das ontologias padrão.

As resposta às questões QP3 e QP4 foram construídas a partir de um esquema de classificação, baseado nas seções de perspectivas futuras dos artigos selecionados. Esse esquema resume, em itens, os potenciais fluxos de pesquisa futuros e limitações apontadas com relação ao tema deste trabalho. Os itens foram utilizados nas respostas a seguir.

QP3) De acordo com os autores, quais as limitações dos trabalhos desenvolvidos sobre abordagem OBDM?

Validação das propostas de abordagem e modelos

A maioria dos autores diz que a proposta apresentada em seu trabalho necessita ampliação de uso ou evoluções (E2 [41], E6 [6], E9 [47], E12 [49] E16 [53], E18 [54], E19 [55], E20 [56], E21 [57], E22 [58], E23 [8]). Isso parece estar relacionado ao número reduzido de estudos, 11, que discorreram sobre a implementação de suas propostas. De mesmo modo, há apenas 5 estudos que envolveram testes/avaliações de suas propostas.

QP4) De acordo com os autores, quais tópicos nas abordagens tipo OBDM necessitam de pesquisa futura ?

***Matching* de ontologias**

Han[45] no estudo E7 e Khan [43] no E4 imaginam que a proposição de métodos e algoritmos de *matching* entre ontologias podem auxiliar na adoção da abordagem ontológica em ambientes nos quais há várias ontologias específicas que necessitam ser integradas. Este tópico é particularmente interessante pois poderia servir para a composição de uma ontologia de Governo a partir de ontologias específicas eventualmente existentes.

Buscas semânticas

Apesar das evoluções nas tecnologias semânticas, parece que ainda não há uma fidedignidade na conversão de buscas a partir de ontologias para *query* em bancos de dados. Lenzerini [31], em E17, considera que as buscas e atualizações semânticas em banco de dados necessitam melhorias de eficiência. Yusuf [48] indica no estudo E11 a necessidade de mais estudos em *semantic query* e indexação.

Anotações semânticas

Li [6], por sua vez, aponta no estudo E6 a necessidade de estudos em anotações semânticas (mapeamento) para conceitos e objetos, pois ainda carecem de formas e instrumentos adequados. [31], ao discorrer sobre o mapeamento de fontes de dados para ontologias, destaca a importância de se documentar todos os dados para fim de governança. Segundo o autor, isso é particularmente importante em grandes organizações, pois as informações sobre os dados são difundidas em documentos separados que são frequentemente difíceis de acessar e raramente estão em conformidade com padrões comuns.

As organizações de grande e médio porte, como a CGU, seriam muito beneficiadas com soluções de mapeamento automático ou semi-automático.

3.7 Discussão dos Resultados

O problema central relatado nos trabalhos selecionados relaciona-se à necessidade de integração de dados de diferentes fontes para permitir a interoperabilidade de sistemas ou para a gestão eficiente e eficaz desse ativo em médias e grandes organizações. A proposição de arquiteturas de informações baseadas em ontologias de domínio específico foi predominantemente utilizada na solução do problema. Para se alcançar a integração desejada, em diversos casos, as propostas foram complementadas com a apresentação de solução técnica que inclui o mapeamento entre fontes de dados e ontologias.

O panorama das pesquisas no tema apresentado na análise preliminar permite concluir que a abordagem ontológica de integração e gestão de dados vem recebendo atenção nos estudos recentes. Indica que a opção por essa abordagem é um caminho a ser considerado para a implementação do processo **PO2** no Ministério CGU. O total de 271 trabalhos/pesquisas identificados desde o ano de 2006 em assuntos que tangenciam temas de governo e a curva de tendência crescente do número de artigos desde aquele ano justificam tal consideração.

A predominância de estudos nesse tema com propostas de modelos ontológicos de dados e arquiteturas técnicas adjacentes aplicados a problemas reais demonstram, de

certa forma, a plausibilidade e aplicabilidade de ontologias nas integrações e gestão de dados, bem como na interoperabilidade entre sistemas. O fato de terem sido descritas 11 experiências com implementação também corrobora a aplicabilidade de tais modelos.

Na questão de pesquisa QP2, foram observados diferentes tipos de ontologia, posteriormente considerados na proposição desenvolvida neste trabalho. As propostas de modelo ontológico dos estudos E1 de Fitzpatrick [2] e E6 de Li [6] foram consideradas particularmente interessantes, pois apresentam grande semelhança de propósito com este trabalho. Tais propostas foram tomadas como referenciais para a proposta desenvolvida para o Ministério CGU, que tem como o seu maior desafio a integração de múltiplos domínios de informação. Além disso, os estudos E1 e E6 citados têm os metadados como seu eixo central, da mesma forma que o processo PO2 do Cobit.

As QP3 e QP4 indicam haver questões em aberto sobre o assunto, porém tais questões são lacunas ou melhorias que não são impeditivas da adoção das ideias contidas nos estudos. A carência de facilidades para mapeamentos entre ontologias ou para anotações (mapeamentos), indexação e *query* semânticas influenciam negativamente na eficiência dos procedimentos, porém, de acordo com Lenzerini [31], uma das vantagens da implementação de OBDM é a possibilidade de extensão do sistema. Pode-se construir um esqueleto do modelo ontológico e se adicionar novos elementos ou fontes de maneira incremental. A compreensão e representação dos domínios, das fontes de dados disponíveis e das relações entre os elementos pode ser paulatina.

3.8 Trabalhos Relacionados

3.8.1 Abordagem Ontológica de Gestão de Dados

A ideia básica da proposta OBDM é uma arquitetura em 3 camadas: 1. Uma ontologia do domínio; 2. Uma camada das fontes de dados e 3. Uma camada de mapeamento entre as duas anteriores (1 e 2). Mais especificamente para a OBDM, a ontologia é uma descrição formal de um domínio de interesse para uma dada organização (ou grupo de usuários), expressa em termos de seus conceitos relevantes, de atributos desses conceitos, de relações entre os conceitos e de afirmações lógicas que caracterizam o conhecimento do domínio. As fontes de dados, por sua vez, são repositórios acessíveis pela organização onde os dados do domínio encontram-se armazenados. Muitas vezes, esses repositórios são numerosos e heterogêneos, sendo cada um gerido e mantido de forma independente dos demais. O nível de mapeamento é uma especificação precisa da correspondência entre dados contidos nas fontes e elementos da ontologia [31].

Essa separação traz algumas vantagens potenciais [32].

- **Primeira vantagem:** A camada ontológica da arquitetura é obviamente o meio que leva a uma abordagem integrativa na governança de dados. Fazendo a representação do domínio de forma explícita, há ganhos em termos de reusabilidade do conhecimento adquirido, o que não ocorre com o uso de esquemas globais (simples unificação de fontes de dados subjacentes).
- **Segunda vantagem:** A camada de mapeamento explicitamente especifica as relações entre conceitos do domínio de um lado e as fontes de dados do outro lado. Tal mapeamento é usado não só para operações dos sistemas de informação, mas também para efeito documental. A importância desse aspecto emerge quando estamos diante de grandes organizações nas quais a informação sobre os dados (metadados) está dispersa em fragmentos de documentação, muitas vezes difíceis de acessar e raramente padronizados. A ontologia e os mapeamentos correspondentes das fontes de dados proveem uma base para a documentação de todos os dados organizacionais, com vantagens óbvias para a governança e a gestão dos sistemas de informação.
- **Terceira vantagem:** Relaciona-se à extensibilidade do sistema. Uma das críticas frequentemente levantada na integração de dados é que ela necessita ser efetuada na origem dos dados, com antecedência, e ela pode ter um custo elevado se implementada posteriormente. No entanto, a abordagem ontológica não impõe a integração total dos dados de uma vez. Em vez disso, até mesmo com um esqueleto básico do domínio, pode se adicionar incrementalmente novas fontes de dados ou novos elementos, quando estiverem disponíveis ou forem necessários, diluindo o custo de integração. Portanto, o *design* geral pode ser considerado como um processo incremental de compreensão e representação do domínio, das fontes de dados disponíveis e das relações entre eles. O objetivo é apoiar a evolução tanto da ontologia quanto dos mapeamentos de tal forma que o sistema continue a operar ao mesmo tempo em que evolui.

A abordagem de gestão de dados proposta neste trabalho utilizou as ideias da OBDM apresentadas em [32, 31, 8] como base. A característica dessa abordagem de separação em camadas é considerada especialmente interessante por facilitar a identificação de impactos mútuos quando da evolução de qualquer dos elementos das camadas.

3.8.2 Mapeamento entre Fontes e Ontologia

O trabalho apresentado por Kohler [10] a respeito de banco de dados semânticos detalha como ocorre o mapeamento entre fontes de dados e ontologia. A principal ideia é mapear tabelas e atributos de um Banco de Dados (BD) para uma determinada ontologia. Essa

ontologia deve ser formal em relação à implementação de uma hierarquia transitiva "is-a", que conecta todos os conceitos. O autor explica os conceitos de semântica de tabela e de atributos, conforme:

A semântica de tabela - pode ser conseguida por meio da ligação da tabela a um ou mais conceitos da ontologia. Isso permite um refinamento no qual todas as entradas, valores, ficam semanticamente anotadas (marcadas) de mesma maneira que a própria tabela.

Uma definição de tabela semântica é uma tupla (T, c) , onde T é uma tabela de banco de dados e $c \in O$, é um conceito da ontologia O .

A semântica de atributos - Os atributos de banco de dados podem ser definidos semanticamente ligando-os a um ou vários conceitos de uma ontologia. Em consequência, os atributos não só podem ser abordados por um conceito mapeado c , mas também pelos sub e super conceitos de c .

Uma definição de atributo semântico é uma tupla (A, c) , onde A é um atributo e $c \in O$, é um conceito da ontologia O .

Tais conceitos de banco de dados semânticos auxiliam, também, no entendimento do propósito de um sistema OBDM que, segundo Daraio [32], é o de prover informações aos usuários para pesquisar dados usando elementos da ontologia para formular as pesquisas. Nesse sentido a OBDM pode ser vista como uma forma de integrar informações, na qual o usual esquema global de dados é substituído por um modelo conceitual de domínio, formulado em uma ontologia expressa em uma linguagem lógica. Com esta abordagem, a visão integrada que o sistema disponibiliza para os usuários não é meramente um modelo que engloba várias fontes de dados, mas sim uma rica descrição semântica dos conceitos relevantes do domínio de interesse, bem como os relacionamentos entre esses conceitos. Essa distinção entre ontologia e fonte de dados reflete a separação entre o nível conceitual apresentado ao usuário e o nível lógico/físico do sistema de informação, que é armazenado nas fontes de dados.

3.8.3 Integração de Dados Baseada em Ontologia

Complementando as ideias OBDM, este estudo considerou a proposta de Fitzpatrick[2], na qual uma abordagem de arquitetura holística e o conceito de dados mestres de múltiplos domínios desempenham um papel central. O autor explica que a ontologia deve distinguir entre o conhecimento de domínio, que pode ser extra organizacional, e o conhecimento localizado no nível de aplicativo. Assim, é coerente com a proposta da OBDM.

Além disso, [2] explica a ideia chamada critério de ortogonalidade, importante para se construir e entender ontologia baseada na arquitetura de referência que ele propõe. O critério de ortogonalidade é definido como o requisito de se basear uma ontologia

recém-criada em uma ou mais ontologias existentes. Segundo o autor, se essa prática for generalizada no desenvolvimento de ontologias, ela ajudaria a reduzir o que chamou de efeito silo [64]. Essa ideia de silo é utilizada também por [8], que ao explicar a arquitetura em silos, diz que ela ocorre quando o sistema é composto por uma infinidade de fontes de dados, independentes e distribuídas, cada uma servindo uma aplicação específica. Esse tipo de arquitetura causa a proliferação descontrolada de dados.

A estrutura de ontologia da arquitetura de referência proposta por Fitzpatrick pode ser visualizada na Figura 3.4. Ela é composta pelas seguintes camadas de cima para baixo: 1) Ontologias de Fundamentação (ou Ontologias de Alto Nível), 2) Ontologia de Integração de Dados Multidomínio, 3) Ontologias de Domínio e Ontologias de Tarefas, no mesmo nível, e 4) Ontologia de Aplicação. O autor explica que à luz do critério da ortogonalidade, a ontologia de integração de dados multidomínio é relacionada, por sub-sunção, a ontologias de fundamentação como a SUMO e outras. As ontologias de domínio são incluídas na ontologia de dados multidomínio. A estrutura compreende, também, ontologias de tarefas genéricas e ontologia de aplicações, esta última para suportar tarefas específicas de domínio.

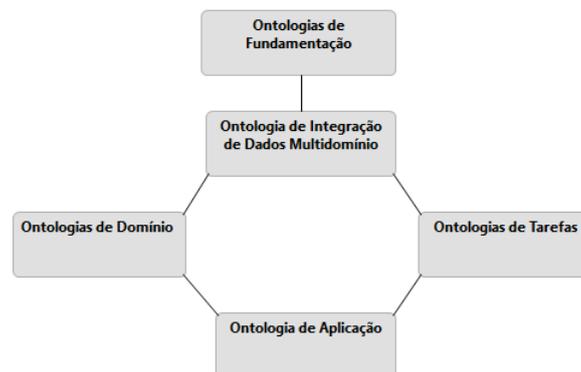


Figura 3.4: Estrutura Ontológica da Arquitetura de Referência [2]

A estrutura proposta por Fitzpatrick [2] permite a implementação do gerenciamento de dados mestre. Ela se encaixa na camada de ontologia da arquitetura OBDM e, por ser holística, permite a uma visão organizada de todos os ativos de dados de uma organização.

Este artigo considera uma composição da ideia central da OBDM[32] e da arquitetura de referência holística de[2] como a base capaz de permitir a integração e o compartilhamento de dados intra e inter-organizacionais.

3.9 Síntese do Capítulo

Este Capítulo apresentou os resultados quantitativos e informações extraídas nas análises sobre os estudos selecionados. A partir deles, foi possível selecionar algumas propostas para servirem de referenciais ao desenvolvimento da solução proposta por este trabalho: proposta de adoção da abordagem ontológica de gestão de dados e modelos de arquitetura da informação e de catálogo (repositório de metadados) coerentes com a abordagem recomendada.

Capítulo 4

Desenvolvimento das Arquiteturas e Estudo de Caso

Neste Capítulo são explicadas as metodologias utilizadas no desenvolvimento das arquiteturas de informação e de solução técnica produzidas neste trabalho. Em seguida, são apresentadas as ideias precedentes relativas ao estudo de caso CGU, tanto do grupo de trabalho PO2 do Órgão quanto de outras iniciativas relacionadas no âmbito do Governo Federal. Por último, os modelos propostos para a CGU são apresentados com os detalhes necessários.

4.1 Metodologias de Desenvolvimento de Arquiteturas

Uma das metodologias que pode ser utilizada no desenvolvimento de arquitetura da informação, na forma de modelo conceitual baseado em ontologia, é a denominada Abordagem de Desenvolvimento de Ontologia "*Enterprise*". Por outro lado, o desenvolvimento de arquitetura de solução técnica, quer para dicionário de dados como para outras necessidades, pode ser apoiado em customização da fase "A. Visão de Arquitetura" do método Togaf ADM - *Architecture Development Method*.

4.1.1 Abordagem *Enterprise* de Desenvolvimento de Ontologia

Há diferentes métodos e abordagens de construção de ontologias descritos no estudo de [15] que poderiam atender as necessidades desta pesquisa, alguns com características em comum. A abordagem *Enterprise* apresenta uma série de características, descritas na Subseção 2.1.3, consideradas interessantes para o estudo deste trabalho e por isso foi

escolhida. A *Enterprise*, como outras metodologias, faz distinção entre duas fases de construção:

- Informal - na qual há uma descrição informal da ontologia com a identificação dos conceitos-chave, definição textual dos conceitos e definição das relações entre eles;
- Formal - na qual a ontologia é incorporada numa linguagem de ontologia.

A customização da fase informal da metodologia *Enterprise* se mostrou suficiente para a intenção deste trabalho, cujo foco é propor modelo de arquitetura da informação com detalhamento suficiente para subsidiar a decisão quanto à abordagem de gestão de dados a ser adotada no estudo de caso. Essa fase permite a criação de esqueleto de modelo de dados baseado em ontologia com os respectivos conceitos-chave, sendo coerente com a extensibilidade do sistema OBDM mencionada por Lenzerini [31].

A primeira etapa da fase informal da abordagem *Enterprise*, denominada Identificação do Propósito, foi guiada pela ideia de desenvolver um modelo conceitual baseado em ontologia cuja estrutura possa ser utilizada para a organização de metadados de fontes de dados. O objetivo buscado é de que a estrutura produzida permita a gestão eficaz e eficiente de dados e informações organizacionais com o suporte para a integração de dados mestres. Além desse objetivo, foram considerados cenários de uso apresentados na Tabela 4.1, adaptada de [65].

A etapa Identificação de Propósito da abordagem *Enterprise* também define o nível de formalidade do modelo. Conforme descrito anteriormente, optou-se pela construção de modelo com base na fase informal da metodologia, na qual a utilização de uma linguagem formal de ontologia é dispensada. Mais especificamente, na classificação quanto ao grau de formalidade de [66], o modelo necessita ter um grau equivalente a uma ontologia "Estruturada informal", ou seja, expressa em linguagem natural, de forma restrita e estruturada. Essa escolha implica que não haverá o processamento do modelo diretamente por máquina, por exemplo em buscas semânticas ou validações automáticas.

Quanto à estrutura, optou-se por uma proposta de modelo de dados formal em relação à implementação de uma hierarquia transitiva "Is-a", que conecta todos os conceitos, conforme a recomendação de [10] para ontologias alvo de mapeamento de fontes de dados. Esse conceito foi expandido incluindo-se hierarquias do tipo "Part-Of".

A segunda etapa da abordagem *Enterprise*, chamada identificação do escopo e que produz a especificação das informações que a ontologia deve caracterizar, foi desenvolvida utilizando questões de competência informais, detalhadas na explicação sobre o modelo de arquitetura da informação proposto.

Por fim, a etapa de "avaliação formal" da abordagem *Enterprise* foi executada sobre o modelo construído em linguagem natural, por meio de verificação do seu atendimento

Tabela 4.1: Propósito do Modelo Baseado em Ontologia

Propósito	Descrição
Intenção de seu uso	A intenção de uso do modelo baseado em ontologia é aplicá-lo na estruturação dos metadados de todas as fontes de dados organizacionais, que possa auxiliar tanto os administrados de dados, em sua função de gestão, quanto os demais usuários na construção e entendimento de modelos de dados.
Possíveis cenários de uso	<ol style="list-style-type: none"> 1. Busca de informações sobre fontes de dados, departamentais ou de aplicações, a partir de conceitos organizados em hierarquia (IS-A). 2. Definição de vocabulário comum entre os atores que produzem ou consomem os dados e informações organizacionais. 3. Guia ao desenvolvimento de aplicações de forma que ocorra o compartilhamento e a integração de fontes de dados. 4. Direcionamento do aprendizado sobre os dados envolvidos nos domínios de negócio da instituição. 5. Busca de informações que permita a identificação de dados mestres e suas fontes.
Usuários finais do modelo baseado em ontologia	<ol style="list-style-type: none"> 1. Administradores de dados; 2. Especialista de <i>business intelligence</i>; 3. Desenvolvedores de aplicações; 4. Especialistas das áreas de negócio que produzem informações gerenciais e estratégicas;

quanto à finalidade e quanto à capacidade de resposta às questões de competência informais. A validação do modelo está descrita mais adiante.

4.1.2 Método ADM - Fase A - para a Arquitetura de Dicionário de Dados

A proposta de arquitetura de solução para o DD visa subsidiar as decisões técnicas futuras sobre a adoção ou não da OBDM para a gestão de dados CGU na implementação do processo PO2 da CGU. A proposta foi elaborada a partir de uma customização da fase 'A. Visão de Arquitetura' do método ADM do Togaf [1], de forma a produzir elementos das saídas: Visão da Arquitetura e Esboço do Documento de Definição de Arquitetura, que foram selecionadas para atender as necessidades do momento. Os conteúdos da Visão de Arquitetura cobertos neste trabalho foram:

1. Cenário de negócios;

2. Descrição do problema;
3. Requisitos principais em alto nível das partes interessadas.

O cenário de negócios e a descrição do problema estão representados pelo relato da Subseção 4.2.1 sobre o problema de pesquisa deste trabalho.

O requisitos principais em alto nível das partes interessadas podem ser resumidos pela descrição dos objetivos de controle PO2.1 e PO2.2 do processo PO2 Definir a Arquitetura da Informação do Cobit 4.1, que também foi a base do trabalho do grupo de trabalho CGU de 2014.

Os elementos do esboço do documento de definição de arquitetura cobertos foram:

1. Arquitetura de dados de linha de base, versão 0.1;
2. Arquitetura de dados alvo, versão 0.1, contemplando conjuntamente os domínios de negócio, tecnologia, dado e aplicação.

A arquitetura de dados de linha de base, versão 0.1, foi representada pelo documento de descrição do Meta-Modelo CGU elaborado pelo grupo PO2. Por outro lado, a arquitetura de dados destino ou alvo, versão 0.1, é a proposta de arquitetura de solução para DD apresentada mais adiante.

Como o objetivo no momento é a criação de uma arquitetura de solução alvo, foram selecionadas os seguintes passos da Fase "A" do método ADM:

- Definição de Escopo;
- Identificação de partes interessadas, seus interesses ou preocupações e requisitos de negócio;
- Desenvolvimento da Visão de Arquitetura.

A definição de escopo da arquitetura de solução DD define o que será coberto no esforço para a arquitetura linha de base. A arquitetura de solução proposta neste trabalho intenciona permitir que diferentes interessados do Órgão possam ter acesso a metadados das fontes disponíveis na CGU, no mínimo no nível de estrutura de dados (por exemplo tabelas de BD, arquivos etc.). A visão criada cobriu, em alto nível, os quatro domínios ToGaf: negócios, dados, aplicações e tecnologia. Particularmente quanto às aplicações de infraestrutura e tecnologias, foi considerada a existência prévia de algumas soluções como conectores para banco de dados, ferramenta de consulta a banco de dados, ferramentas de ETL e ferramenta de criação de ontologias.

A identificação de partes interessadas para a elaboração da arquitetura alvo deste trabalho ficou restrita aos atores mais importantes que atuam na solução proposta. A

proposta de arquitetura considerou os usuários finais da solução, já enumerados como usuários finais do modelo na tabela 4.1, e as partes interessadas responsáveis pela gestão e manutenção do modelo. A gestão do modelo foi considerada de responsabilidade dos administradores de dados, com interesse particular nos entes ou coisas transversais à organização. Os especialistas de negócio e desenvolvedores de aplicações foram considerados fontes importantes para a manutenção das informações sobre os domínios de negócio CGU e sobre as aplicações e serviços e tarefas gerais comuns, respectivamente. Esses interessados podem agregar valor às suas tarefas cotidianas ao utilizar e disponibilizar uma visão comum dos dados da CGU.

Uma das técnicas recomendadas pelo método ADM para descobrir e documentar requisitos de negócios é a utilização de cenários de negócio. Os possíveis cenários de uso do modelo que se encontram listados na tabela 4.1 foram utilizados, também, como referencial para o desenvolvimento da arquitetura de solução DD.

A arquitetura de solução desenvolvida foi documentada por meio de diagrama produzido na ferramenta Archi [67] com a respectiva discriminação dos seus elementos, apresentados na seção 4.4.

4.2 Contexto do Estudo de Caso

4.2.1 O Meta-Modelo PO2 da CGU

O modelo conceitual de metadados da CGU, arquitetura alvo, foi desenvolvido pelo grupo de trabalho formado em 2014 sob o nome Meta-Modelo da Arquitetura da Informação. Segundo a documentação produzida [3], os componentes da Arquitetura da Informação da CGU podem ser visualizados por meio da Figura 4.1. Pelas cores na Figura, percebe-se que o Meta-modelo foi dividido em três grandes grupos: o Grupo das Entidades, no centro do diagrama, grupo do Dicionário de Dados, à esquerda, e o Grupo da Documentação, elementos em branco à direita.

A Entidade, elemento central do modelo, refere-se a elementos computacionais que guardam informações relevantes para o negócio, cujos metadados serão carregados automaticamente no dicionário de dados por meio de um processo de carga.

Uma Entidade é dividida em três níveis de especialização. Cada nível representa uma granularidade específica:

- **Repositório** - representa um banco de dados ou um arquivo que possua informações relevantes ao negócio e que deva estar classificado no dicionário de dados.

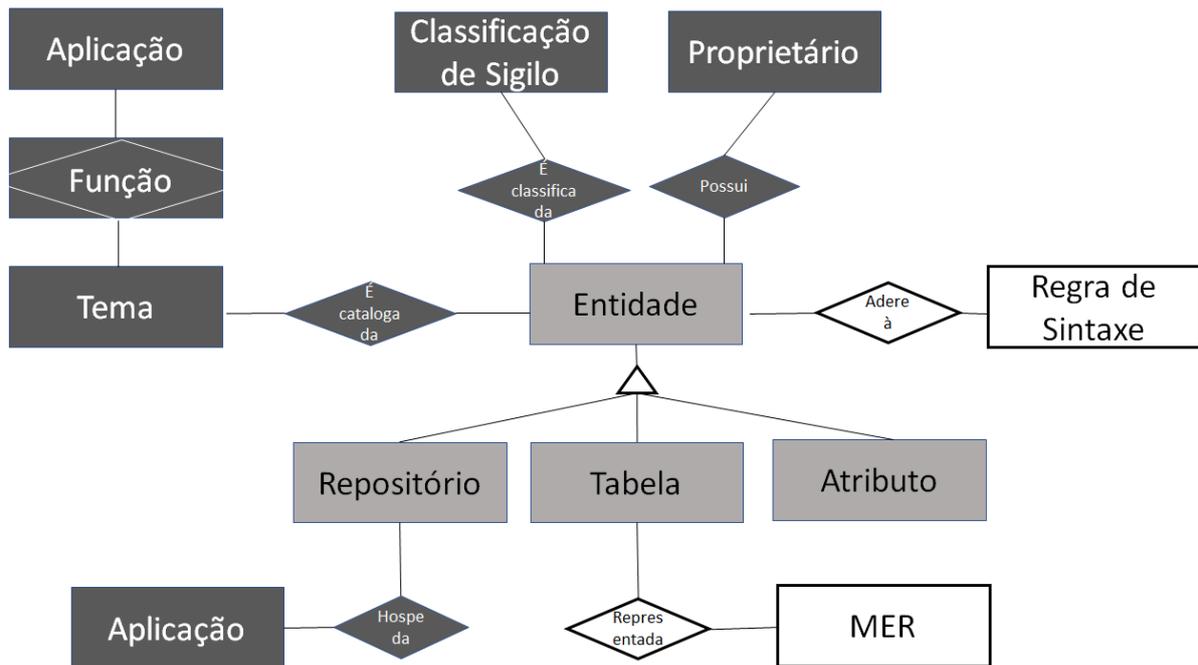


Figura 4.1: Meta-modelo de Arquitetura da Informação [3]

- **Tabelas** - dizem respeito à relação de colunas e linhas utilizadas em bancos de dados relacionais. Dentro do Meta-Modelo da Arquitetura da Informação, as Tabelas devem fazer parte de um Modelo Entidade Relacionamento – MER.
- **Atributos** - representam as colunas de uma determinada tabela dentro de um banco de dados.

Toda a contextualização da informação deve estar armazenada em um dicionário de dados. Assim, todos os componentes do grupo Dicionário de Dados, na cor escura, devem ser cadastrados no Dicionário de dados da Arquitetura da Informação.

- **Classificação de Sigilo** - define a qualificação de um determinado elemento de dados CGU no que diz respeito ao seu sigilo. Existe Portaria desde 2012 regulamentando o assunto no Órgão.
- **Proprietário** - Unidade organizacional responsável por uma determinada Entidade que esteja cadastrada no dicionário de dados.
- **Tema** - lista de temas negociais com os quais as aplicações estão relacionadas e suas respectivas entidades. O Tema é o elo responsável por conectar as diversas aplicações da casa às suas respectivas entidades conforme o tema negocial atendido. Exemplo:

– 1.1 Controle Interno

- 1.1.1 Controle Interno - Avaliação da Gestão dos Administradores (AGA);
- 1.1.2 Controle Interno - Avaliação da Execução de Programas do Governo (AEPG).

...

- **Aplicação** - existem dois conceitos diferentes para as aplicações. O primeiro diz respeito às aplicações que estão armazenadas em diferentes repositórios (Banco de Dados) e o outro diz respeito à principal aplicação responsável por atualizar uma determinada tabela em um Banco de Dados.

Os componentes da Arquitetura da Informação também podem estar documentados na *Wikipedia* do Órgão ou em outras ferramentas, como é o caso dos componentes do Grupo Documentativo, MER e Regra de Sintaxe. Esses componentes têm a função de direcionar o desenvolvimento de aplicações, que devem estar alinhadas às definições da Arquitetura da Informação.

- **Regra de Sintaxe** - definem os padrões a serem seguidos na criação de objetos no Banco de Dados.
- **Modelos de Entidade-Relacionamento (MER)** - são documentos auxiliares que permitem uma visualização das tabelas de um determinado sistema e seus relacionamentos.

O modelo proposto pelo grupo de trabalho **PO2 CGU** englobou os aspectos técnicos e de gerenciamento das estruturas de dados do Órgão, porém, os aspectos semântico-negociais não foram abordados na ocasião, à exceção dos componentes Função-Tema. Em razão disso, foi referenciado neste trabalho como um modelo conceitual essencialmente de metadados técnicos. Ele ainda carece de elementos como conceitos e suas relações para facilitar o entendimento comum dos dados, indispensável à indução do uso e compartilhamento otimizado da informação.

O modelo foi elaborado no esforço inicial do grupo de trabalho **PO2 CGU**, porém sua implementação ainda não ocorreu. Assim, complementações ou ajustes ainda são oportunos, até mesmo para aderência a iniciativas gerais do Governo Federal. Nesse sentido, foram examinados alguns produtos da Administração Pública Federal (APF) sobre o tema, os quais são apresentados na Seção 4.2.2.

4.2.2 Iniciativas no Governo Federal

A busca da sistematização e conhecimento dos dados na Administração Pública Federal - APF¹ - não é recente, por isso, foram examinadas três de suas iniciativas de forma a

¹A APF corresponde ao conjunto de órgãos da administração direta, autárquica e fundacional.

inspirar ou servir de insumo para uma proposta de abordagem e arquitetura da informação da organização - AI - para o Ministério CGU:

1. **O Modelo de Referência de Dados - (MRD)** – é um dos documentos do Framework de Arquitetura Corporativa para Interoperabilidade no Apoio à Governança - FACIN[4], em elaboração na Administração Pública Federal;
2. **O e-VoG - Vocabulários e Ontologias do Governo Eletrônico** – é um conjunto de padrões, ferramentas e metodologias para possibilitar o intercâmbio de informações com acordo semântico, com uso de ontologias como ferramenta para explicitar conhecimentos de maneira formal. Um dos produtos do e-VoG é o Repositório de Vocabulários e Ontologias de Governo Eletrônico, local para acesso a todas as referências ontológicas do Governo Eletrônico Federal. Nesse repositório encontra-se o Vocabulário Controlado do Governo Eletrônico (VCGE) [68], que é um vocabulário controlado para indexar informações no governo federal.
3. **Modelo Global de Dados - (MGD)** – modelo de dados e processos em alto nível e tem como objetivo mapear as informações armazenadas nos sistemas do Governo Federal. Este modelo foi adotado pela e-PING [69] e definido como a Arquitetura de Interoperabilidade para Integração de Dados e Processos[70].

Modelo de Referência de Dados do FACIN

Desde meados de 2014, o Segmento Áreas de Integração para o Governo Eletrônico da Arquitetura ePING de Interoperabilidade desenvolveu, com participação de representantes da sociedade (cidadãos, governos, organizações e empresas), ações no sentido de permitir a implementação de um modelo de referência, que possa ser adotado em nível Federal, Estadual ou Municipal, para promover uma melhoria constante na Governança de Dados, Processos, Tecnologia e Organizações. Como referência para os trabalhos iniciais, foram utilizados os padrões de mercado em arquitetura corporativa *Togaf* e *Zachman Framework*. Este último é um padrão proprietário disponibilizado por *Zachman International, Inc* [71], sendo um dos mais antigos e utilizado por empresas e governos ao redor do mundo desde meados dos anos 80.

O *Framework* de Arquitetura Corporativa para Interoperabilidade no Apoio à Governança - FACIN[72], resultado do trabalho iniciado pelo segmento de integração do e-PING, tem por objetivo construir uma visão consistente dos modelos de representação das capacidades de cada organização governamental e servir como uma referência para os diversos órgãos da APF. Espera-se que o estabelecimento de Arquitetura Corporativa e de padrões de interoperabilidade apoiará a Estratégia de Governança Digital Brasileira, am-

pliando a colaboração entre as organizações do Governo Federal e melhorando a eficiência dos serviços de governo eletrônico para a sociedade.

O *framework* define modelo de referência como uma fonte autorizada de informações sobre uma área específica que orienta a uma utilização homogênea e integrada das visões de várias arquiteturas e soluções. Os modelos e padrões de referência servem como uma base para arquiteturas e soluções e podem também ser utilizados para fins de comparação e de alinhamento. Existem 9 diferentes modelos de referências para suportar o desenvolvimento das visões e respectivos artefatos ao longo do uso do FACIN, dentre eles está o Modelo de Referência de Dados - MRD, que dá suporte à visão de dados e informações. O MRD [4] apresenta, às organizações governamentais, a forma como os dados e informações devem ser descritos e relacionados entre si, além de estabelecer padrões e diretrizes para sua descrição, categorização, compartilhamento, reuso e gestão. Os elementos definidos, a relação entre eles e os artefatos, as diretrizes e os padrões associados descrevem o MRD, que estabelece princípios comuns acerca da implementação e da gestão de informações.

O MRD oferece suporte, entre outros, para o desenvolvimento de Modelo de Dados Lógicos, Dicionário de Dados e de Modelo Conceitual de Dados (que descreve e categoriza o conjunto de dados relevantes ao governo e o relacionamento entre eles). O escopo do MRD abrange [4]:

1. **Modelo Conceitual de Referência de Dados** que contém os meta-conceitos e seus relacionamentos necessários à definição, descrição, classificação e entendimento dos dados e informações e dos relacionamentos entre elas;
2. **Taxonomia** que descreve a categorização e hierarquia de dados e informações classificadas por área de domínio para ser utilizada por todo o governo e as organizações governamentais;
3. **Conjunto de Documentos Comuns à Visão de Dados**;
4. **Conjunto de Políticas e Diretrizes** para implementação e gestão do MRD;
5. **Conjunto de Padrões e Modelos de Referência** de interesse de todas as organizações governamentais relacionadas ao MRD.

Os parágrafos seguintes descrevem o Modelo Conceitual em 1 e a Taxonomia de Referência citada em 2 por terem relação mais estreita com as saídas Dicionário de Dados Corporativos e Modelo de Arquitetura da Informação do Cobit.

Segundo o MRD [4], **O Modelo Conceitual de Referência de Dados**, diagrama apresentado na Figura 4.2, agrupa metadados em três áreas. A área de **Descrição dos dados**, contornada com pontilhados, provê os metadados para descrever a estrutura sintática e semântica dos dados. Permite a comparação entre dados de organizações governa-

mentais. A área de **Contexto dos dados**, que contém o Ator de Negócio e Papel, provê a categorização dos dados e informações e uma taxonomia. A padronização do contexto dos dados estabelece uma abordagem para uniformização da categorização dos ativos de dados e, dessa maneira, estabelece as bases para a Governança de dados no Governo como um todo. A área de **Compartilhamento de dados**, composta por Ativo de Dados, Serviço de Dados e Definição do Corpo de Dados provê os metadados necessários à definição de formas de acesso e trânsito de dados entre organizações governamentais.

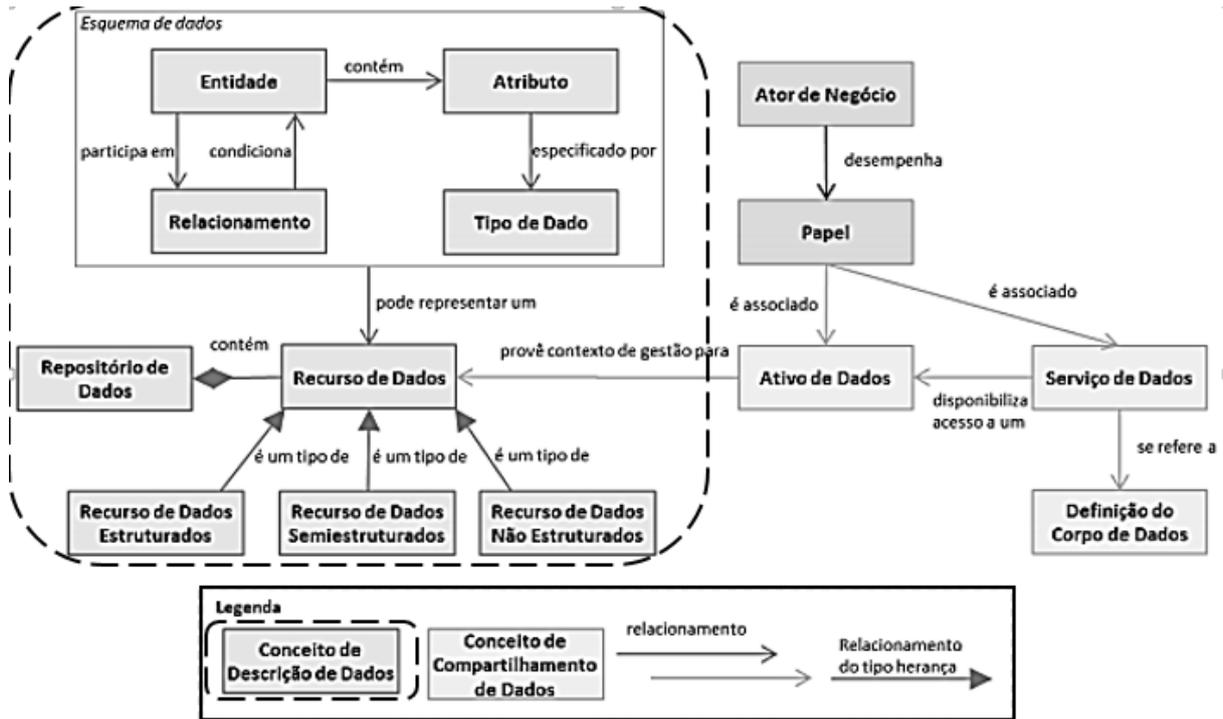


Figura 4.2: FACIN - Modelo Conceitual de Referência de Dados [4]

De modo geral, os componentes da área de Descrição de Dados, pontilhada, correspondem aos metadados técnicos que compõem o grupo das Entidades do **Meta-Modelo de AI CGU**, mas a Descrição de Dados MRD vai um pouco além, pois inclui os conceitos dos dados.

A área de Compartilhamento de Dados não encontra correspondentes no Meta-Modelo, contudo não se considera isso uma restrição do modelo CGU, pois o escopo inicial a ser abrangido não inclui o compartilhamento de dados entre organizações.

A novidade se apresentada na área de Contexto Dados, na qual há a presença dos conceitos de Taxonomia e de Categoria de Informação, elementos semânticos do modelo. Segundo o MRD, a Taxonomia representa um vocabulário de termos disposto em uma estrutura hierárquica que provê a categorização ou classificação da informação. Uma Taxonomia contém uma ou mais Categorias de Informação, as quais proveem contexto a

um dado dentro de uma Taxonomia. O mapeamento entre as fontes de dados estruturados e a respectiva semântica (taxonomia/categoria de informação) se dá pelo conceito de Ativos de Dados, que representa um conjunto de dados que podem estar em algum formato como um banco de dados relacional, um sítio eletrônico, um documento, serviço de dados, etc.

O Modelo de Conteúdo do FACIN recomenda a utilização do Vocabulário Controlado de Governo Eletrônico (VCGE) como base para a construção das taxonomias. Como será visto mais adiante, os elementos Função/Tema do Meta-Modelo CGU apresentam certa afinidade com a estrutura em árvore do VCGE, permitindo uma análise futura de composição entre essas estruturas.

Há um trecho do MRD que trata dos fatores críticos de sucesso na implementação do modelo, dentre eles são citados:

- Definição da soberania da informação (dados-mestres);
- Criação e manutenção da taxonomia;
- Estabelecimento de uma visão comum de todos os repositórios utilizados no armazenamento de dados e informações.

O modelo **MRD-FACIN** englobou aspectos técnicos, gerenciais e alguns semânticos (definição, vocabulário e taxonomia) sobre dados do Governo. Não há menção ao uso de modelagem conceitual de dados baseada em ontologias como forma de melhor induzir ao uso e compartilhamento otimizado da informação.

Vocabulários e Ontologias do Governo Eletrônico - e-VoG

O e-VoG é um conjunto de padrões, ferramentas e metodologias para possibilitar: o intercâmbio de informações com acordo semântico, de forma a viabilizar o pronto cruzamento de dados de diversas fontes; o uso de metodologias de modelagem conceitual como forma de eliciação do conhecimento tácito das áreas de negócio de governo; o uso de ontologias como ferramenta para explicitar conhecimentos de maneira formal e coerente; o alinhamento conceitual das diversas áreas do conhecimento do governo. Um dos produtos do e-VoG é o Repositório de Vocabulários e Ontologias de Governo Eletrônico, local para acesso a todas as referências ontológicas do Governo Eletrônico Federal. Nesse repositório encontra-se o Vocabulário Controlado do Governo Eletrônico (VCGE) [68].

O documento descritivo do VCGE explica que ele é um vocabulário controlado para indexar informações no governo federal, projetado com dois objetivos básicos: 1) interface de comunicação com o cidadão e 2) ferramenta de gestão. Como ferramenta de gestão ele deve ajudar os gestores a gerenciarem suas informações. A expectativa é que o VCGE

seja usado para classificar qualquer conteúdo de informação (documentos, bases de dados, mídia eletrônica, documentos em papel, etc) que não seja classificado outra forma mais específica de indexação. O VCGE é feito para ser consultado pelo público geral e seu processo de indexação deve ser feito por pessoas que não são profissionais especializados (biblioteconomistas, arquivologistas, etc).



Figura 4.3: Representação dos Termos do VCGE em Forma de Árvore

O modelo CGU não contempla vocabulário controlado e ontologias. Os Temas/Função definidos na CGU não se encontram relacionados ao VCGE, mas podem ser futuramente incluídos nesse vocabulário.

Modelo Global de Dados

O Serviço Federal de Processamento de Dados - Serpro, desenvolveu o Modelo Global de Dados - MGD a partir do Projeto de Integração (PRINT) iniciado em 2008. O objetivo foi construir um modelo que viabilizasse e garantisse a integrabilidade das informações geradas no âmbito do Macroprocesso de Planejamento, Orçamento e Finanças - MPOF, para suportar adequadamente as necessidades de informações inerentes ao processo decisório da APF [70].

Tal iniciativa resultou no desenvolvimento das partes componentes do MGD: metodologia, padrões de notação e modelo de governança. Além disso, a iniciativa produziu o documento Modelo Conceitual e Metodológico para apresentar: o histórico e o contexto de criação do Projeto; o diagnóstico realizado pelo Comitê Interministerial do Macroprocesso de Planejamento, Orçamento e Finanças; a proposta de solução que deu origem ao Modelo Global de Dados.

Segundo o Modelo Conceitual e Metodológico [73], para a elaboração do modelo foi utilizada a modelagem essencial de dados e processos - MED, que se chama Essencial porque considera apenas as informações relevantes para o entendimento do negócio, não se atendo aos detalhes operacionais ou tecnológicos. Ela tem como objetivo representar esquematicamente os dados tratados pelos sistemas informatizados que compõem o Macroprocesso de Planejamento, Orçamento e Finanças, visando criar uma plataforma de metadados integrados. A MED atua, portanto, de forma semelhante a um processo de engenharia reversa, resgatando o modelo de dados a partir das fontes de informação:

- Descrição da base de dados, ou metadados, dos sistemas legados;
- Percepção dos dados pelos analistas de negócio responsáveis pelos sistemas legados;
- Percepção dos dados pelos usuários finais desses sistemas.

Os produtos finais elaborados a partir da modelagem de dados e processos realizada em cada participante do Macroprocesso são consolidados em um documento chamado Compêndio. Ele reúne os vários artefatos produzidos sob a metodologia MGD, dentre os quais os mais relacionados à semântica dos dados e ao entendimento do processo de negócio são: Diagrama de Processo de Negócio; Dicionário de Dados; Matriz de Entidades versus Processos de Negócio e Diagrama de Integração Completo. Este último agrupa todos os diagramas de entidade-relacionamento em uma mesma visão. As integrações são representadas na forma de elementos gráficos. Exemplos de alguns pontos de integração identificados: Órgão SIORG, Empenho, Localidade, Servidor/Outros; Fornecedor, Material/Serviço, Ação e Contrato.

- **Diagrama de Processo de Negócio** - representação gráfica do Processo de Negócio, baseada na notação BPMN;
- **Diagramas de Entidades-Relacionamentos de alto nível** - contendo somente as informações essenciais ao negócio, que são identificadas tendo como foco a integração, podendo agrupar várias entidades da vida real em apenas uma entidade do modelo;
- **Dicionário de Dados** - onde todas as entidades serão registradas e definidas. Ele é um complemento essencial do Diagrama de Entidade-Relacionamento, estabelecendo uma semântica precisa para as informações do Macroprocesso;
- **Matriz de Entidades versus Processos de Negócio** - provê o cruzamento das Entidades existentes no MGD e os Processos de Negócio, especificando em quais Processos cada Entidade é gerada, consumida e atualizada;
- **Diagrama de Integração Completo** - disposição de todos os diagramas de entidade-relacionamento agrupados em uma mesma visão. É nele que ficam representadas as integrações na forma de elementos gráficos, tanto para as integrações existentes como para as propostas de integração. Permite identificar as integrações entre as soluções de cada Ministério mapeado e as integrações entre todas as soluções mapeadas. Exemplos de alguns pontos de integração identificados: Órgão SIORG, Empenho, Localidade, Servidor/Outros; Fornecedor, Material/Serviço, Ação e Contrato.

O Meta-Modelo da **AI CGU** não prevê diagramas de processo de negócio nem modelos entidade-relacionamento de alto nível agrupados. No entanto, aquele meta-modelo indica o relacionamento existente entre os componentes Entidade e Tema (que de certa forma espelha as áreas de atuação da CGU). O dicionário de dados da CGU não prevê o fornecimento de definições semânticas para as informações, previstas no dicionário MGD. A visão holística fornecida pelo Diagrama de Integração Completo MGD não tem paralelo no modelo CGU, particularmente porque esse novo modelo definido pelo MGD traz destaque às entidades que são transversais a vários sistemas.

As iniciativas da APF citadas apresentam alguns elementos semântico-negociais não presentes no modelo CGU como: definições; vocabulário; taxonomia; modelo global de dados e integrações; taxonomia e ontologia, mesmo que incipiente. A identificação da existência desses elementos não significa que eles devam ser incorporados em modelo a ser proposto, necessitam ser avaliados um a um para verificar a adequação à realidade da CGU, inclusive sob o aspecto de importância para os objetivos a serem alcançados.

Assim, as iniciativas APF citadas, ainda que em desenvolvimento e evolução, podem trazer insumos para o estabelecimento do processo PO2 ou da prática de gestão APO03.02

nos Órgãos. Vale ressaltar, que não foi identificado, entre as iniciativas, modelo com estrutura que atenda às necessidades CGU de ter uma arquitetura da informação com características de visão abrangente de dados e informações e de viés para o compartilhamento e a integração de dados.

4.3 Modelo de Arquitetura da Informação

A Arquitetura da Informação (AI) organizacional deste trabalho foi proposta sob a forma de um protótipo de modelo conceitual de dados suportado por ontologia. Como o propósito e escopo do modelo proposto são abrangentes e considerando a extensibilidade do sistema OBDM descrita por [32], nesta etapa o esqueleto do modelo e os respectivos conceitos-chave não constituirão uma ontologia propriamente dita, ou seja, um modelo desenvolvido a partir de uma conceituação compartilhada por todos os interessados [11].

De toda maneira, este esqueleto de modelo e os respectivos conceitos-chave deverão atender às questões de competência informais listadas a seguir:

- Dado um ente ou coisa do mundo real, por exemplo pessoa no papel de Servidor da CGU (interno), é possível saber 'Quantas e quais são as possíveis fontes de dados que armazenam dados pessoais de servidores ?'
- Dado um ente do mundo real, por exemplo instituições de Governo ou ligadas a ele, é possível saber Quantas e quais são as possíveis fontes de dados relativas às estruturas organizacionais de governo, independentemente da esfera ou de poder ?
- Dada uma tabela corporativa, por exemplo Órgão SIORG, é possível saber "Quais são as aplicações que a utilizam ?"
- Dado um ente ou coisa do mundo real, por exemplo organização pública, é possível saber 'Quais são as aplicações que mantêm dados sobre o ente organização pública ?'
- Dado um ente do mundo real, por exemplo o próprio Órgão CGU, é possível saber 'Quais as possíveis fontes de dados mantidas por aplicações próprias que mantêm cadastro da estrutura CGU e que poderiam ser integradas ? '
- Dada uma organização ligada ao Governo Federal, por exemplo a estatal Petrobras, é possível saber 'Quais são as aplicações e as respectivas estruturas de dados que devem ser consideradas para se obter informações sobre todas as ações de prevenção da corrupção relacionadas à estatal ?'

- Dada uma base de dados departamental com dados pessoais de pessoas estrangeiras que não residem no Brasil, seria possível 'Incorporar (mapear) as tabelas dessa base para elemento(s) da estrutura de AI proposta ?'

Sob o ponto de vista do Modelo de Conteúdo Togaf, a proposta de AI deste trabalho é composta de blocos de construção do tipo Entidades de Dados (nível conceitual de dados) relativos a blocos pertencentes à Arquitetura de Negócios e à Arquitetura de Sistemas. Então, por exemplo, no modelo de AI proposto para a CGU há entidades de dados relativas a Organização, Local, Eventos e Funções pertencentes ao conjunto Arquitetura de Negócios do modelo Togaf. Haverá, também, entidades de dados relativas a Serviços de Sistemas de Informação e Aplicação que fazem parte do conjunto Arquitetura de Sistemas.

Estrutura do Modelo Conceitual de Dados Holístico

A proposta deste trabalho para a AI CGU, Figura 4.4, baseada na arquitetura de referência de ontologia [2], é, de mesma forma que essa referência, um modelo conceitual holístico que compreende as camadas: Camada de Integração de Dados Multidomínio - MD; Camada de Domínios de Negócio - DN; Camada de Tarefas Gerais - TG; Camada de Aplicações - AP.

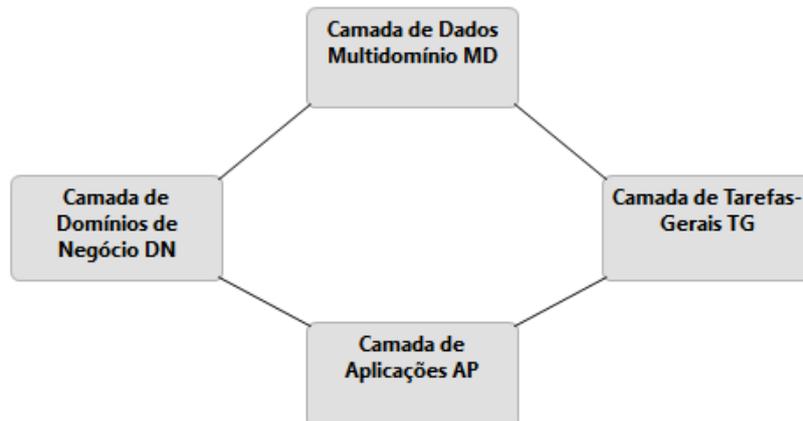


Figura 4.4: Estrutura do modelo baseado em ontologia

Os próximos parágrafos detalham os conteúdos de cada camada e os relacionamentos existentes entre os domínios de dados que as compõem. As Figuras 4.5 a 4.8 foram reproduzidas do protótipo de modelo conceitual de dados suportado por ontologia (AI CGU) construído na ferramenta WebProtégé [74]. Cada camada com os seus respectivos domínios de dados são mostrados na ferramenta como uma classe e suas subclasses organizadas em hierarquia.

Camada de Dados Multidomínio MD

A principal finalidade da camada de integração de dados multidomínio MD é conter os conceitos fundamentais, bem conhecidos na área de modelagem de dados, que permitam a integração e compartilhamento de dados e a consequente interoperabilidade de sistemas. Isso significa que esta camada pode auxiliar as respostas a questões como **Q1** e **Q2** em 1.1, que envolvem dados mestres.



Figura 4.5: Classes da Camada de Integração MD

A Figura 4.5 identifica as classes que compõem a camada de integração MD. Elas foram selecionadas principalmente de ontologias organizacionais utilizadas como referência e acrescidas de determinados conceitos gerais de forma a constituir uma hierarquia de domínios de dados abrangente. Algumas referências utilizadas foram, entre outras: Modelo Conceitual de Ontologia Organizacional em [75], Metamodelos de Conteúdo da Visão de Negócios [76], o Modelo de Conteúdo Togaf [1], domínios de dados identificados para a Ontologia Multidomínio de Dados de Integração de Fitzpatrick [2]. Mesmo essas referências não sejam próprias do domínio governamental, elas puderam ser utilizadas em razão da generalidade dos conceitos multidomínio.

Esta camada não tem elemento similar no Meta-Modelo CGU e da mesma forma, não foi identificada estrutura dedicada a organizar dados mestres no MGD e nem nos modelos disponibilizados pelo FACIN. Esses dois últimos recomendam a utilização de rótulos para destacar os de dados de integração ou dados-mestres da organização.

Camada de Domínios de Negócio DN

As finalidades primordiais da camada de negócios DN são: mostrar quais são os domínios de negócio CGU e explicar a essência de cada um deles a partir da descrição dos relacionamentos entre classes pertencentes à camada MD. A camada DN, ao explicar as funções desempenhas pela CGU sob a ótica de domínio de dados, pode auxiliar as respostas a questões no estilo **Q3** da subseção 1.1.

Complementarmente, a camada DN mostra algumas classes que são específicas de um domínio de negócio, por exemplo: a classe relativa à "Autorização para Exercício de Atividade", do domínio "Prevenção de Conflito de Interesses". Essas classes específicas são interessantes para compor uma explicação mais pormenorizada do domínio de negócio. Além disso, elas pertencem, também, à hierarquia formada na camada MD, permitindo a sua definição mais precisa.

A Figura 4.6 mostra, em seu lado esquerdo, algumas classes que compõem a camada de negócio DN. Elas representam funções afetas à CGU, com suas divisões, que foram derivadas das informações constantes em [7]. Além disso, a figura mostra algumas classes específicas de domínio de negócio DN.

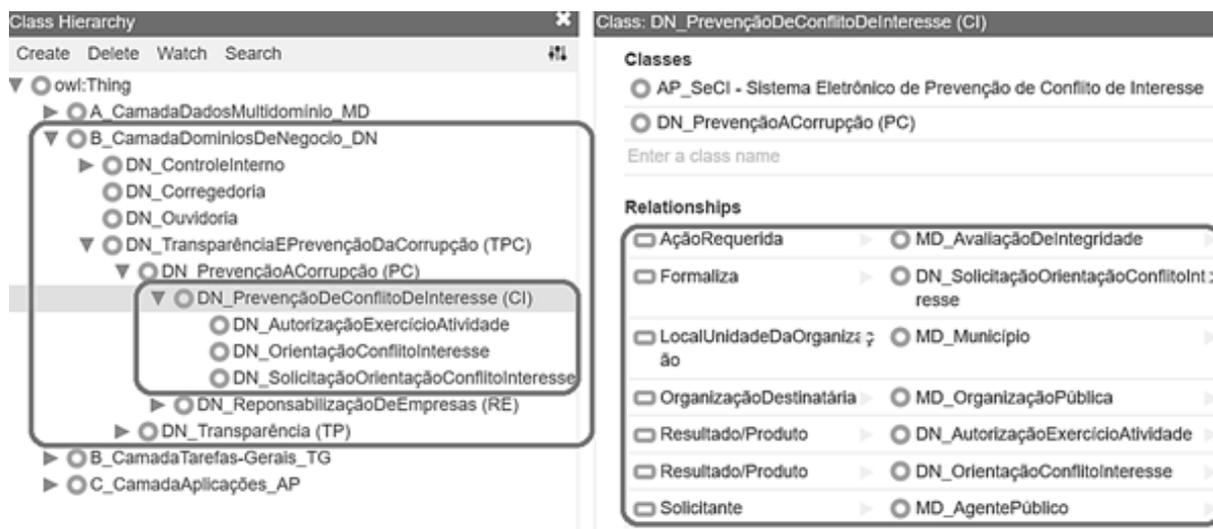


Figura 4.6: Classes da Camada de Domínios de Negócio DN

A Figura 4.6 mostra, em seu lado direito, como algumas classes se inter-relacionam a fim de descrever uma das funções CGU (Prevenção de Conflito de Interesse) na seguinte forma geral: Sujeito (Ex:Partes Envolvidas) + Ação (Ex:Tipo de Ação do Órgão) + Objeto (Ex:Item de Recurso) + Complementos (Ex:Local ou Documento). Todas as classes necessárias para descrever cada função CGU nesse formato geral têm de ser encontradas na camada de integração MD ou na própria camada de domínios de negócio DN. O fato dessa camada demonstrar somente domínios de dados de negócio e seus inter-relacionamentos

a torna consistente com a visão GDM, que é produzida a partir de modelagem Essencial de sistemas. Por outro lado, ao ser organizada pelas funções CGU, ela fornece uma visão similar à fornecida pelos elementos Função/Tema do Meta-Modelo CGU.

Camada de Tarefas-Gerais TG

A finalidade da camada TG é revelar os serviços ou tarefas de suporte gerais que fazem parte de aplicações CGU e que são independentes de domínio de negócio. São serviços ou tarefas de natureza corporativa e que podem ser reusados.

A Figura 4.7 identifica algumas classes que compõem a camada de tarefas e serviços gerais TG, que é organizada por assuntos. Os assuntos listados são meramente ilustrativos, podendo ser evoluídos na medida em que as aplicações CGU forem documentadas. Esta camada pode dar suporte ao reuso de serviços e ao compartilhamento de dados como, por exemplo, no caso da tarefa de concessão de acesso e perfil para aplicações CGU que encontra-se sob o assunto "Segurança".

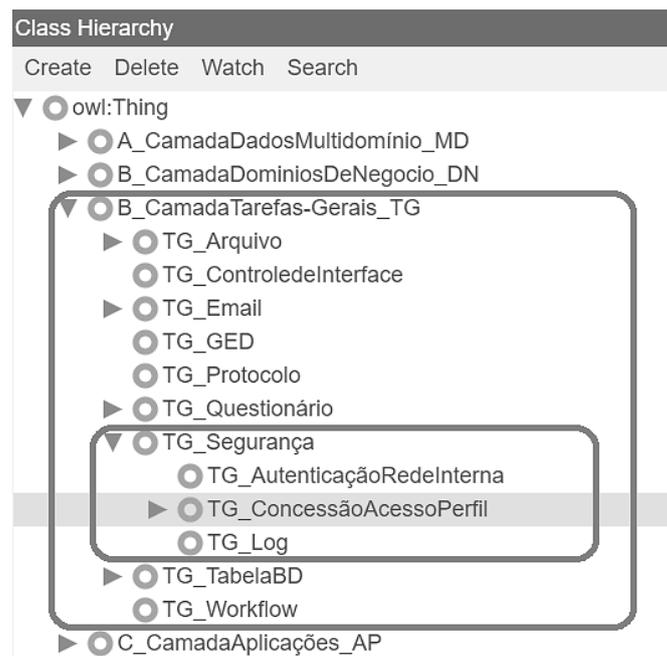


Figura 4.7: Classes da Camada de Tarefas ou Serviços Gerais TG

Esta camada não tem elemento similar no Meta-Modelo CGU.

Camada de Aplicações AP

Por último, a camada de aplicações AP tem como principal finalidade mostrar quais são as aplicações CGU existentes e qual a sua composição sob a perspectiva de negócio e

de serviços/tarefas corporativos. Assim, a composição de cada aplicação é demonstrada basicamente por meio de classes provenientes da camada DN e da camada TG.

Complementarmente, a camada AP mostra algumas classes que são específicas de uma aplicação. Apesar de específicas, essas classes devem ser, também, subclasse de alguma outra pertencente à camada MD, DN ou TG.

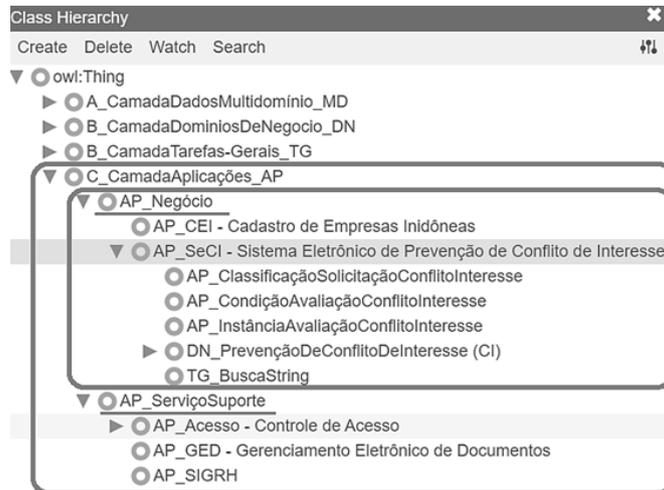


Figura 4.8: Classes da Camada de Aplicações AP

A camada AP está organizada em duas divisões: Negócio e Serviço de Suporte. As aplicações do Órgão mais concernentes às funções do Órgão ficarão sob a primeira divisão e as mais concernentes ao oferecimento de serviços de suporte às atividades finalísticas do Órgão ficarão sob a última divisão. A Figura 4.8 apresenta alguns exemplos de aplicações com a classificação devida.

A visão desta camada é semelhante à fornecida pelo elemento Aplicação do Meta-Modelo CGU.

O modelo conceitual de dados baseado em ontologia proposto para a CGU, de abordagem *top-down*, foi organizado em camadas, diferentemente de modelos no estilo do GDM recomendado pelo FACIN. Este último é criado a partir de uma composição de modelos ER de aplicações que interagem entre si, numa abordagem *bottom-up*. Apesar da diferença entre abordagens, a abordagem GDM poderia ser utilizada para verificar a completude e exatidão do modelo proposto, particularmente no que diz respeito à identificação dos pontos de integração entre aplicações. Outra forma de verificar o nível de atendimento do modelo proposto é por meio do mapeamento das fontes de dados das aplicações CGU existentes para classes dos modelo proposto, funcionando como uma engenharia reversa. Assim, para que o modelo seja considerado completo e correto é necessário que cada fonte de dados encontre uma classe ou relacionamento correspondente em qualquer das camadas (MD, DN, TG ou AP) do modelo proposto.

4.4 Arquitetura de Solução de Dicionário de Dados

A proposta de arquitetura de solução de dicionário de dados (DD) CGU foi orientada pela ideia base da OBDM [32] de separação entre a ontologia e as fontes de dados, com o mapeamento entre cada fonte e classes da ontologia. Ela é uma arquitetura alvo versão 0.1 e tem como objetivo auxiliar a decisão sobre a aprovação da solução proposta e o entendimento dos seus impactos, por isso englobou elementos dos quatro domínios Togaf (negócio, aplicação, dados e tecnologia).

O desenho da arquitetura, Figura 4.9, foi elaborado na ferramenta livre Archi², que implementa a especificação da linguagem Archimate[77]. As descrições dos principais elementos da arquitetura encontram-se na tabela 4.2.

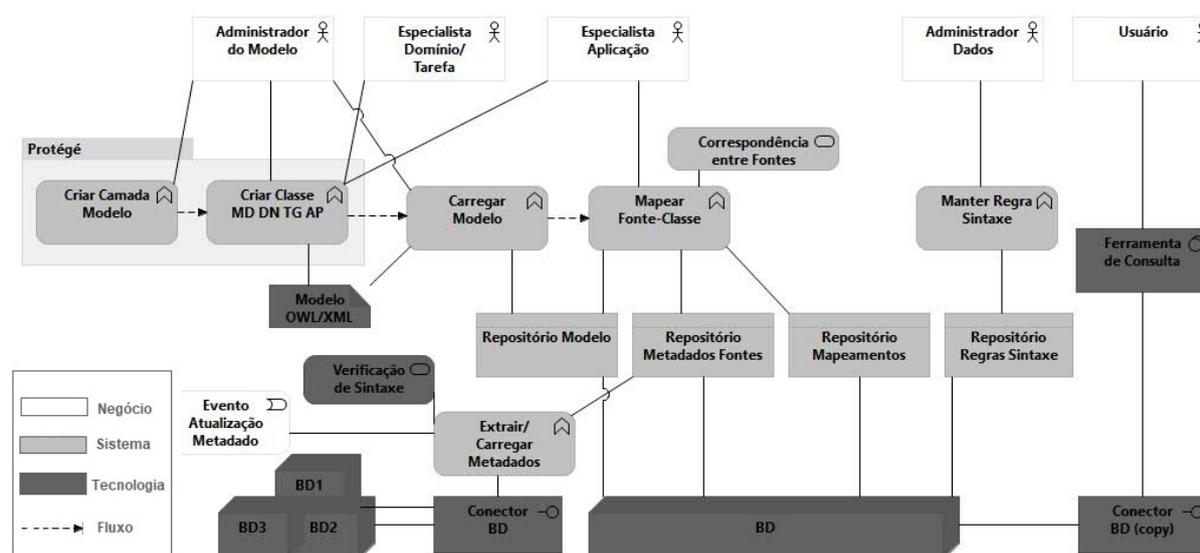


Figura 4.9: Arquitetura de Solução de DD

A solução de DD intenciona, mais especificamente, atender os requisitos do processo PO2.2 Dicionário de Dados Corporativo e Regra de Sintaxe de Dados do Cobit. O suporte oferecido pela solução a cada requisito do objetivo de controle são explicados a seguir:

- Compartilhamento de dados entre aplicações e sistemas - disponibiliza metadados a respeito das fontes de dados CGU de uma forma organizada por domínios de dados, sobretudo a respeito dos dados transversais existentes na organização, ou seja, candidatos a dados mestres;
- Entendimento comum dos dados entre a área de TI e os usuários de negócio - disponibiliza conceitualização sobre as classes de dados em linguagem natural. Isso é

²Ferramenta livre de modelagem Archimate. Download disponível em <https://www.archimatetool.com/>

Tabela 4.2: Elementos da Arquitetura de Solução de DD

Elemento:	Descrição:
Administrador de Dados	Ator responsável pela manutenção das regras de sintaxe para a criação de estruturas de dados.
Administrador do Modelo	Ator responsável pela manutenção da estrutura básica do modelo (camadas e as suas principais hierarquias) e por auxiliar os outros mantenedores de classes do modelo a mantê-lo coerente.
Especialista de Aplicação	Ator que é especialista em determinada aplicação e que é responsável pela sua descrição.
Especialista de Domínio/Tarefa	Ator que é especialista em determinado domínio, tarefa ou serviço e que é responsável pela sua descrição.
Usuário	Ator que utiliza as informações disponíveis no DD.
Criar Camada do Modelo	Função que permite a criação da estrutura básica do modelo (camadas e principais hierarquias).
Criar Classe MD DN TG AP	Função que permite a criação de classe: de multidomínio ou dados mestres MD; de domínio de negócio DN; de tarefa ou serviço-geral TG; de aplicação AP. Deve permitir, também, a criação de relacionamentos entre classes.
Carregar Modelo	Função que permite a carga, em repositório relacional, de dados do modelo construído em ferramenta de construção de ontologias. A função deve permitir que o administrador efetue ajustes no modelo em repositório.
Extrair/Carregar Metadados	Função que permite a extração automática de metadados técnicos referentes às fontes de dados estruturadas e a carga dos metadados no repositório respectivo. A carga dos metadados pode incluir o resultado da verificação automática da sintaxe das estruturas de dados a serem carregadas.
Manter Regra de Sintaxe	Função que permite o mapeamento entre metadados das fontes de dados estruturados (tabelas) e as classes do modelo. Ela pode incluir soluções automáticas de correspondência entre tabelas para auxiliar o trabalho de mapeamento.
Mapear Fonte-Classe	Função que permite o mapeamento entre metadados das fontes de dados estruturados (tabelas) e as classes do modelo. Ela pode incluir soluções automáticas de correspondência entre tabelas para auxiliar o trabalho de mapeamento.

particularmente relevante para o entendimento da essência dos domínios de negócio do Órgão, que está presente na hierarquia de classes da camada DN do modelo conceitual baseado em ontologia.

- Evitar a criação de elementos de dados incompatíveis - permite o exame das fontes de dados existentes sobre um mesmo domínio de dados. Há facilidade para identificar todas as fontes de dados de uma determinada classe, bem como os relacionamentos

existentes entre essa classe e outras;

- Incorporação das regras de sintaxe da organização - permite o armazenamento das regras de sintaxe em repositório específico e a consequente criação de verificações automáticas da observância dessas regras nas estruturas de dados criadas.

Em estágios futuros de experimentação poderá ocorrer a incorporação, a simplificação ou até mesmo a eliminação de alguns blocos da solução. Por exemplo, a remoção do bloco responsável pela extração e carga automática do modelo conceitual produzido em ferramenta de ontologia para o repositório de modelo conceitual.

Capítulo 5

Validação e Análise

A proposta de Arquitetura da Informação (AI), formulada por meio de protótipo de modelo conceitual de dados baseado em ontologia, foi verificada primeiramente a partir do mapeamento de tabelas pertencentes a dois esquemas em banco de dados da CGU. Em seguida, foi avaliada por um questionário respondido por especialistas do Órgão. Essas duas atividades de validação são descritas nas próximas seções.

As atividades de validação e análise deste estudo não contemplaram a arquitetura de solução de DD por não existir modelo paralelo no trabalho original do grupo PO2 CGU. Ademais, o objetivo dessa arquitetura é auxiliar a decisão sobre qual será o modelo de arquitetura de informação a ser utilizado na CGU e, se escolhido modelo baseado em ontologia, quais são os seus impactos.

5.1 Mapeamento entre Fontes e Classes

De acordo com a proposta apresentada na Seção 4.4, o mapeamento entre fontes e classes do modelo conceitual deve ocorrer em solução que implemente essa função e as demais indicadas naquela arquitetura de DD. Como tal solução deve ser desenvolvida no futuro, em tempo de implantação do processo PO2 CGU, a experimentação ocorrida neste trabalho utilizou a própria ferramenta WebProtégé simulando tal funcionalidade.

A simulação dos mapeamentos entre classes e fontes de dados na ferramenta de ontologia foi efetuada a partir das seguintes escolhas:

- Utilização da aba "*Individuals*" da ferramenta WebProtégé para conter as tabelas a serem mapeadas. *Individuals* podem referir-se a objetos ou processos, são entidades que não são tipos [78], ou seja, a aba comporta as instâncias de classes;
- Mapeamento no nível de tabela. O mapeamento pode ocorrer no nível de tabela ou de atributo [10]. Ao mapear uma tabela para um conceito é atribuída uma semântica

à própria tabela, bem como para todos os seus valores. Dessa forma, mesmo que a simulação não contemple o mapeamento no nível de atributos, podemos considerá-los como se fossem valores associados à tabela, estendendo-lhes a marcação recebida pela tabela;

- Mapeamento de cada tabela somente para o seu principal conceito ou classe. Uma mesma tabela pode conter conjuntos de atributos relacionados a diferentes conceitos. Por exemplo, a tabela corp.Servidor contém atributos específicos do servidor (nome, data de nascimento etc.), mas também contém atributos de endereço do Órgão de origem do servidor. Para efeito de simplificação, será escolhido somente o conceito que abrange os principais atributos da tabela, no caso da tabela corp.Servidor houve mapeamento para a classe MD ServidorInterno (servidores da CGU) ;
- Escolha de fontes de dados pertencentes a uma aplicação que atenda a alguma das funções CGU e das fontes sob o esquema de bases corporativas do Órgão. As bases corporativas são interessantes para este experimento por serem candidatas naturais a dados mestres.

A Tabela 5.1 apresenta as informações gerais sobre as fontes de dados escolhidas:

Tabela 5.1: Resumo das Fontes de Dados Mapeadas

Esquema	Nr.Tabelas	Observação
.corp	46	Este esquema contém tabelas que são utilizadas por diversos sistemas CGU. Elas são provenientes de aplicações do próprio Órgão ou provenientes de sistemas estruturantes do Governo Federal .
.seci	37	Este esquema contém as tabelas que atendem ao Sistema de Prevenção de Conflito de Interesses - SeCI desenvolvido na CGU. Há tabelas mantidas pelas funcionalidades da própria aplicação e tabelas que são cópias de tabelas corporativas.
Total	83	

5.1.1 Resultados dos Mapeamentos

As observações sobre os resultados, alguns esperados, dos mapeamentos efetuados foram:

- Todas as tabelas do esquema corp. foram mapeadas para classes pertencentes somente à camada de dados multidomínio MD. Não foi necessária a criação de nenhuma nova classe de primeiro nível da hierarquia da camada para contemplá-las;
- O esquema SeCI teve 12 (doze) tabelas mapeadas para classes eminentemente MD. Há 9 (nove) cópias de tabelas corporativas e 3 (três) da própria aplicação SeCI;

- O esquema SeCI teve 9 (nove) tabelas mapeadas para a camada de domínio de negócios, todas para uma única classe: DN SolicitaçãoOrientaçãoConflitoInteresse. Tal subclasse foi criada por opção para dar maior clareza sobre o tipo de solicitação tratado pelo SeCI. As nove tabelas armazenam informações sobre as solicitações SeCI e encontram-se em estruturas distintas para atender a regras de modelagem de dados;
- O esquema SeCI teve 06 (seis) tabelas mapeadas para classes TG. A engenharia reversa das tabelas restou por identificar uma nova subclasse, BuscaString, que foi acrescentada à camada e que diz respeito a tarefa/serviço que pode ser reutilizado por outras aplicações;
- O esquema SeCI teve 10 (dez) tabelas mapeadas para classes AP, que foram criadas para abarcar as especificidades da aplicação, de interesse somente neste nível.

O mapeamento de tabelas requer a análise de sua estrutura e, algumas vezes, a análise de seus valores. Além disso, é recomendado que seja efetuado por especialista da aplicação responsável pelas tabelas por haver peculiaridades que não são esclarecidas pela análise citada.

5.2 Avaliação do Modelo Conceitual de Dados

A validação do modelo conceitual necessita ser referendada pelos potenciais envolvidos em sua manutenção e pelos seus principais usuários. Assim, foram definidos os seguintes objetivos para essa etapa de avaliação do modelo proposto:

Objetivo 1 : Verificar se o modelo atende às questões de competência que guiaram a sua elaboração;

Objetivo 2 : Verificar a percepção de melhorias do modelo proposto neste trabalho com relação ao Meta-Modelo elaborado pelo grupo de trabalho PO2 CGU.

A segunda etapa de validação foi efetuada por avaliação qualitativa. Segundo [79], a adoção de métodos qualitativos de pesquisa se justifica quando, por exemplo, o estudo do sistema ocorre em seu contexto real ou em estudos de fenômenos complexos, nos quais os fatores contextuais devem ser analisados. Essas justificativas se encaixam bem neste estudo de caso de problema real de gestão de dados da CGU.

Segundo [80], as fontes de dados qualitativos podem ser: a observação; a observação em campo; entrevistas e questionários; documentos e textos; impressões e reações do

pesquisador. Desse modo, é cabível a aplicação de questionário a especialistas CGU como meio de avaliação da proposta de modelo conceitual deste trabalho.

Antes da aplicação do questionário, porém, foi necessária a apresentação do modelo proposto por se tratar de uma estrutura nova e com elementos de nível de abstração elevado. Esse fato explica a necessidade de restringir a pesquisa aos atores mais importantes. As apresentações, ocorridas para pequenos grupos de especialistas, serviram como uma espécie de pré avaliação do modelo, pois no decorrer decorrer das explicações foram efetuados questionamentos sobre entendimento do conteúdo do modelo, bem surgiram questões práticas de uso.

A primeira aprovação do questionário efetuada foi a de conteúdo, por meio de pré-teste junto a especialista semelhante ao público escolhido [81]. Essa atividade gerou refinamentos de conteúdo para melhorar a clareza e compreensão do questionário. Além disso, ocasionou a criação de espaço no questionário para registro de avaliações pessoais, no qual o avaliador pôde anotar sua apreciação pessoal.

A amostra com 10 (dez) participantes foi escolhida de forma não probabilística, por escolha razoável, visando escolher uma amostra representativa do público-alvo, estando de acordo com as orientações propostas por [81]. Os participantes têm atuação nas seguintes especialidades: administração de banco de dados (**AD**); desenvolvimento de aplicações (**DES**); *business intelligence* (**BI**), que produzem informações gerenciais do Órgão; produção de informações gerenciais e estratégicas (**INF**), por meio de cruzamento e mineração de dados; assessoramento em TI (**ASS**). Há especialistas que atuam em mais de uma especialidade. A tabela 5.2 mostra um panorama das áreas de atuação dos especialistas.

Tabela 5.2: Participantes da Avaliação

ID do Especialista	Especialidade(s)	Função	Tempo na(s) especialidade(s)
E1	AD; BI	Analista da Equipe de AD	De 5 a 10 anos
E2	INF	Analista Área Negócio	Mais de 10 anos
E3	AD	Analista da Equipe de AD	De 5 a 10 anos
E4	AD	Analista da Equipe de AD	Até 5 anos
E5	AD	Analista da Equipe de AD	De 5 a 10 anos
E6	INF; BI; ASS	Assessor TI	De 5 a 10 anos
E7	AD; BI; INF	Gerente da Equipe AD	Mais de 10 anos
E8	DES	Analista Desenvolvedor	Mais de 10 anos
E9	INF	Analista Área Negócio	Mais de 10 anos
E10	INF	Analista Área Negócio	De 5 a 10 anos

5.2.1 Perguntas do Questionário

O questionário foi dividido em duas seções: a primeira, com 8 perguntas, foi praticamente formada pelas questões de competência utilizadas para a elaboração do modelo conceitual de dados; a segunda, com 4 perguntas, faz questionamentos sobre a percepção de melhorias trazidas pelo modelo proposto neste trabalho em relação ao Meta-Modelo CGU existente.

As perguntas definidas para a primeira seção encontram-se na Tabela 5.3.

Tabela 5.3: Questionário - Perguntas da Primeira Seção

Pergunta
Q1 Dada uma tabela corporativa, por exemplo Órgão do Siorg ¹ , é possível saber "Quantas são as possíveis replicações dessa tabela, quer sejam totais ou parciais ?"
Q2 Dada uma tabela corporativa, por exemplo Órgão do Siorg, é possível saber "Quais são as aplicações que a utilizam ?"
Q3 Dado um ente ou coisa do mundo real, por exemplo pessoa no papel de Servidor da CGU (interno), é possível saber "Quantas e quais são as possíveis fontes de dados que armazenam dados pessoais de servidores ?"
Q4 Dado um ente ou coisa do mundo real, por exemplo organização pública, é possível saber "Quais são as aplicações que mantêm tais dados ?"
Q5 Dada uma base de dados departamental CGU com dados pessoais de pessoas estrangeiras que não residem no Brasil, seria possível "Incorporar (mapear) as tabelas dessa base para elemento(s) da estrutura de AI proposta ?"
Q6 Dado um ente do mundo real, por exemplo instituições de Governo ou ligadas a ele, é possível saber "quantas e quais são as possíveis fontes de dados relativas às estruturas organizacionais de governo, independentemente da esfera ou de poder ?"
Q7 Dado um ente do mundo real, por exemplo o próprio Órgão CGU, é possível saber "Quais as possíveis fontes de dados mantidas por aplicações próprias que mantêm cadastro da estrutura CGU e que poderiam ser integradas ? "
Q8 Dada uma organização ligada ao Governo Federal, por exemplo a estatal Petrobras, é possível saber "Quais são as aplicações que devem ser consideradas para se obter informações sobre todas as ações de prevenção da corrupção relacionadas à estatal ? "

As questões definidas para a segunda seção encontram-se na Tabela 5.4.

Os participantes tiveram acesso liberado para consultas ao protótipo do modelo conceitual proposto bem como ao protótipo do Meta-Modelo CGU disponíveis na ferramenta WebProtégé. Essa ferramenta de ontologia é de fácil uso, mas ainda possui uma curva de aprendizado na qual há uma dificuldade inicial de aprender as possibilidades de configuração para as suas consultas e para a descoberta dos diferentes caminhos para chegar a um mesma informação. Por isso, cada questão foi precedida de uma imagem com resultado de consulta relacionada à questão apresentada.

Tabela 5.4: Questionário - Perguntas da Segunda Seção

Pergunta	
Q9	O modelo baseado em ontologia é capaz de melhorar a identificação dos dados mestres da CGU? (dados mestres representam os objetos de negócios que são comuns a toda a organização e por isso são compartilhados.)
Q10	O modelo baseado em ontologia é capaz de melhorar o entendimento de cada domínio de negócio da CGU ?
Q11	O modelo baseado em ontologia é capaz de melhorar o entendimento do papel dos entes/coisas nos diferentes domínios de negócio CGU ?
Q12	O mapeamento de tabelas de uma base de dados departamental, por ex. com dados restritos sobre os agentes públicos, para o modelo baseado em ontologia pode melhorar "o entendimento e uso dessas tabelas, inclusive por usuários não especialistas de TI ?

5.2.2 Resultados da Avaliação

Resultados da Primeira Seção do Questionário

Devido à natureza dos questionamentos para a avaliação do modelo de AI proposto, as escalas definidas para o questionário foram, segundo definição de [82], com critério de importância subjetivo e de valores não métricos (que são em níveis, não são em valores numéricos).

Segundo [83] na tomada de decisão por qual escala utilizar, deve-se observar o objetivo da pesquisa e os possíveis entrevistados. Assim, foram definidos dois objetivos para a coleta de percepção dos especialistas na primeira seção do questionário. O principal objetivo é saber o grau de resposta do modelo conceitual proposto a cada questão de competência (não é possível responder; a resposta é parcial; a resposta é completa). O objetivo complementar é saber se a obtenção de resposta é fácil ou difícil.

Segundo [84], [85], [86] os quatro tipos ou classes de escala mais comuns são: nominal, ordinal, intervalar e proporcional (ou de razão). Para atender aos dois objetivos simultaneamente, foi gerada uma composição de tipos: ordinal e nominal. As escalas ordinais distinguem os diferentes valores da variável, hierarquizando e classificando de acordo com diferentes graus, estabelecendo uma evolução, sem, no entanto, quantificar a magnitude da diferença face aos outros indivíduos [85, 86]. As escalas nominais, por outro lado, são classificativas permitindo descrever as variáveis ou designar os sujeitos, sem recurso à quantificação [85].

A escala foi composta para atender ao principal objetivo de avaliar o grau de resposta do modelo, por escala ordinal, e o complementar de saber se a resposta é fácil ou difícil, por escala nominal. A escala adotada foi:

0 - não é possível responder;

- 1 - a resposta é parcial e difícil;
- 2 - a resposta é parcial, mas fácil;
- 3 - a resposta é completa, mas difícil;
- 4 - a resposta é completa e fácil.

O resumo dos resultados relacionados à primeira seção, por pergunta (questão de competência), é apresentado na Figura 5.1.

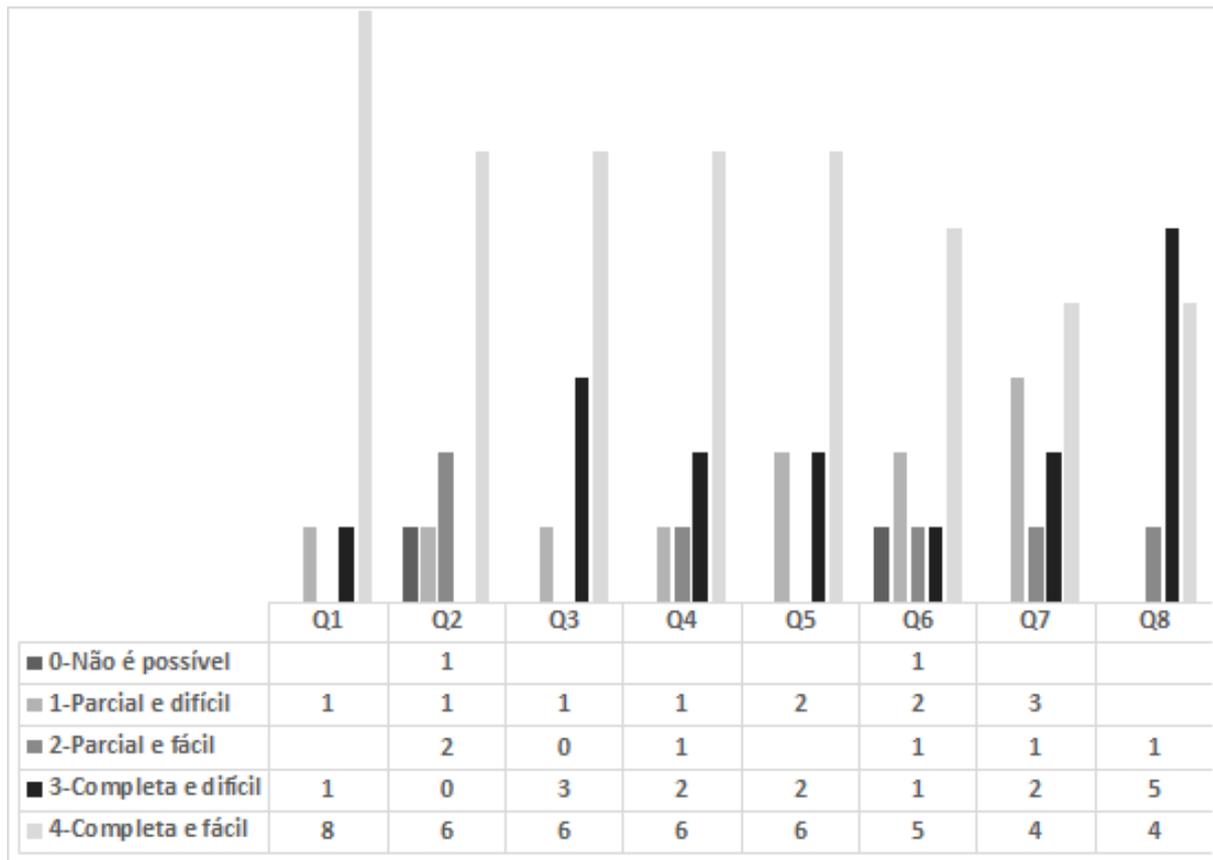


Figura 5.1: Distribuição das Manifestações por Pergunta da Primeira Seção

Cada pergunta corresponde a um total de 10 manifestações, correspondentes ao total de especialistas avaliadores. As perguntas Q1 a Q6 serão analisadas em pares, formando universos de 20 manifestações.

As questões de competência Q1 e Q2 indagam sobre a capacidade de obtenção de metadados de uma fonte de dados já conhecida, ou seja, são consultas cuja resposta se inicia em exame na camada OBDM de fontes de dados. A maioria (14 em 20 manifestações) demonstra que a percepção predominante é de que o modelo fornece RESPOSTA COMPLETA e FÁCIL a essas questões.

As questões Q3 e Q4 indagam sobre a capacidade de identificação das fontes de dados relacionadas a uma classe selecionada a partir de sua conceitualização. Trata-se de consultas que se iniciam em exploração das hierarquias e classes que formam as camadas conceituais do modelo. Quase todos os participantes (17 em 20 manifestações) demonstra que a percepção é de que o modelo é capaz de dar RESPOSTA COMPLETA a essas questões. Porém, 5 (cinco) dessas manifestações são no sentido de que, apesar de fornecer respostas completas, elas são DIFICEIS de se obter.

As questões de competência Q5 e Q6 são sutilmente diferentes das questões Q3 e Q4, pois indagam sobre situações em que há conceitos bem específicos e que não se encaixam devidamente nas classes existentes. Há duas possibilidades de solução: usar conceitos mais gerais já existentes ou criar os conceitos requeridos para a situação encaixando-os nas hierarquias devidas. Em ambos os casos, a solução inicia-se pela exploração das hierarquias e classes existentes de modo a compreendê-las. Um número expressivo de participantes (14 em 20 manifestações) demonstra que a percepção é de que o modelo é capaz de dar RESPOSTA COMPLETA a essas questões. Porém, 3 (sete) dessas manifestações são no sentido de que, apesar de fornecer respostas completas, elas são DIFICEIS de se obter. Além disso, 6 manifestações indicaram que a percepção é de que o modelo dá RESPOSTA PARCIAL ou NÃO RESPONDE a essas questões. Este último fato pode ser consequência de os especialistas não terem entendido a característica de extensibilidade do modelo.

A Q7 indaga sobre a identificação de fontes de dados que poderiam ser integradas por se referirem a um mesmo ente do mundo real. O texto da pergunta foi precedido de figura mostrando a lista de todas as fontes de dados mapeadas para o ente da questão. Houve a 6 (seis) percepções de que o modelo RESPONDE à questão, porém 2 (duas) delas acham a resposta DIFÍCIL. Observa-se que nesta questão houve um número proporcionalmente maior de manifestações com RESPOSTA PARCIAL (4 em 10 manifestações) quando comparada as outras questões da seção. Isso era um resultado esperado, em razão de o protótipo não trazer detalhes no nível de colunas das tabelas de banco de dados mapeadas. Isso demonstra a necessidade de o dicionário de dados contemplar metadados sobre as colunas das tabelas, porém, não implica a necessidade de haver mapeamento nesse nível de detalhe.

A questão Q8 também indaga sobre a integração de dados, porém sob a perspectiva de compartilhamento de dados para compor informações gerenciais sobre um domínio de negócio como um todo. É uma questão voltada à compreensão da camada de negócios DN do modelo conceitual. Há 9 (nove) manifestações com RESPOSTA COMPLETA, porém a maioria delas percebe ser DIFÍCIL a obtenção das respostas.

A escala de resposta para as perguntas da primeira Seção é única, permitindo aglutinar todas as manifestações a essas perguntas num único gráfico para resumir a percepção geral

da capacidade de o modelo responder às questões de competência. A Figura 5.2 apresenta o resumo do número total de manifestações por item da escala.

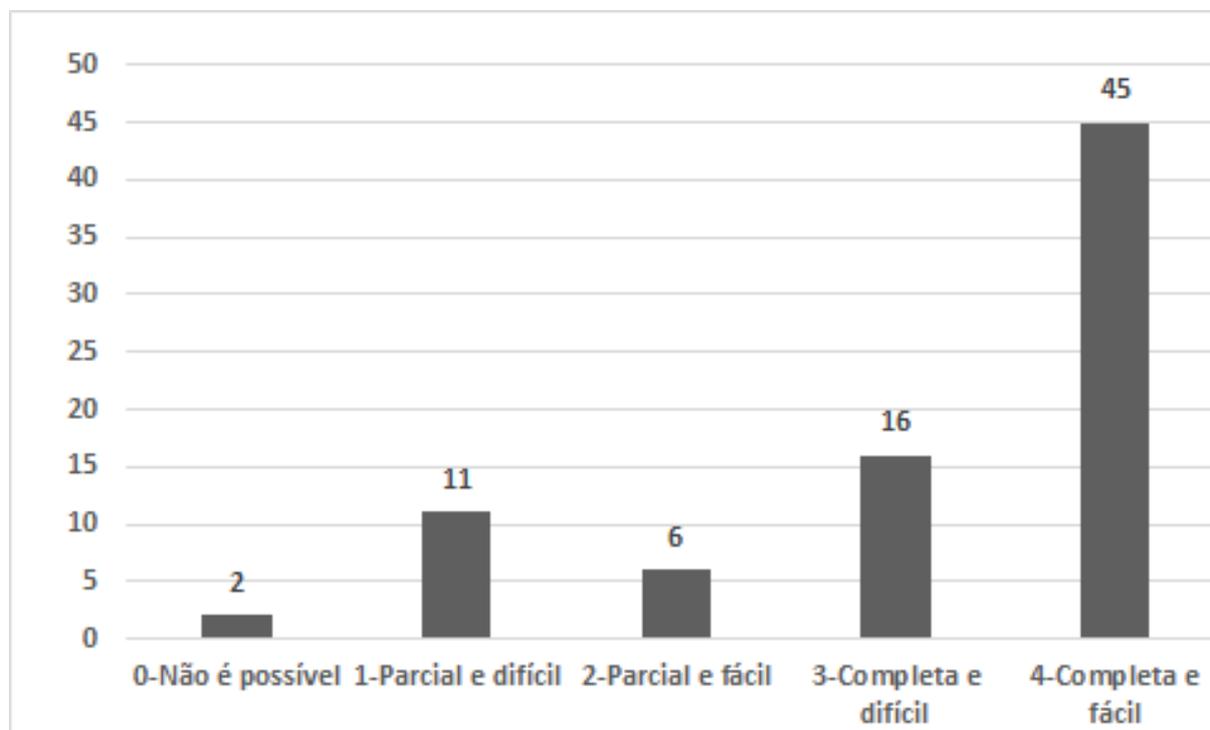


Figura 5.2: Resumo das Manifestações às Questões da Primeira Seção

De um modo geral o modelo foi considerado capaz de responder às questões de competência, em um universo de 80 itens respondidos respostas há 61 marcações para os itens 4 e 3 - o modelo dá RESPOSTA COMPLETA. Porém nesse universo de resposta COMPLETA há 16 marcações para o item 3, ou seja, a resposta é DIFÍCIL de se obter.

Houve uma manifestação pessoal sobre as questões da seção 1: "... facilita o trabalho de identificação e contextualização da informação, porém o maior desafio está na criação da cultura de uso da AI como ferramenta de trabalho e na manutenção de modelos complexos".

Resultados da Segunda Seção do Questionário

A segunda seção do questionário também teve dois objetivos para a coleta de percepção dos especialistas. O principal objetivo é apurar o grau de melhoria das capacidades, sob diferentes aspectos, do modelo conceitual proposto quando comparado ao Meta-Modelo CGU. Os graus considerados, em escala ordinal, foram: não há melhoria, há pouca melhoria e melhora bastante. O objetivo complementar é saber se o aspecto avaliado é

importante ou não, em escala nominal, para o especialista. A composição de objetivos gerou os seguintes itens:

- 0 - não há melhoria;
- 1 - há pouca melhoria e este aspecto não é importante;
- 2 - melhora bastante, mas não é importante;
- 3 - há pouca melhoria, porém é importante;
- 4 - melhora bastante e este aspecto é importante.

Os dados das manifestações às perguntas da segunda seção do questionário, que objetiva verificar a percepção dos especialistas quanto a possíveis melhorias trazidas pelo modelo conceitual de dados proposto neste trabalho, encontra-se apresentado na Figura 5.3.

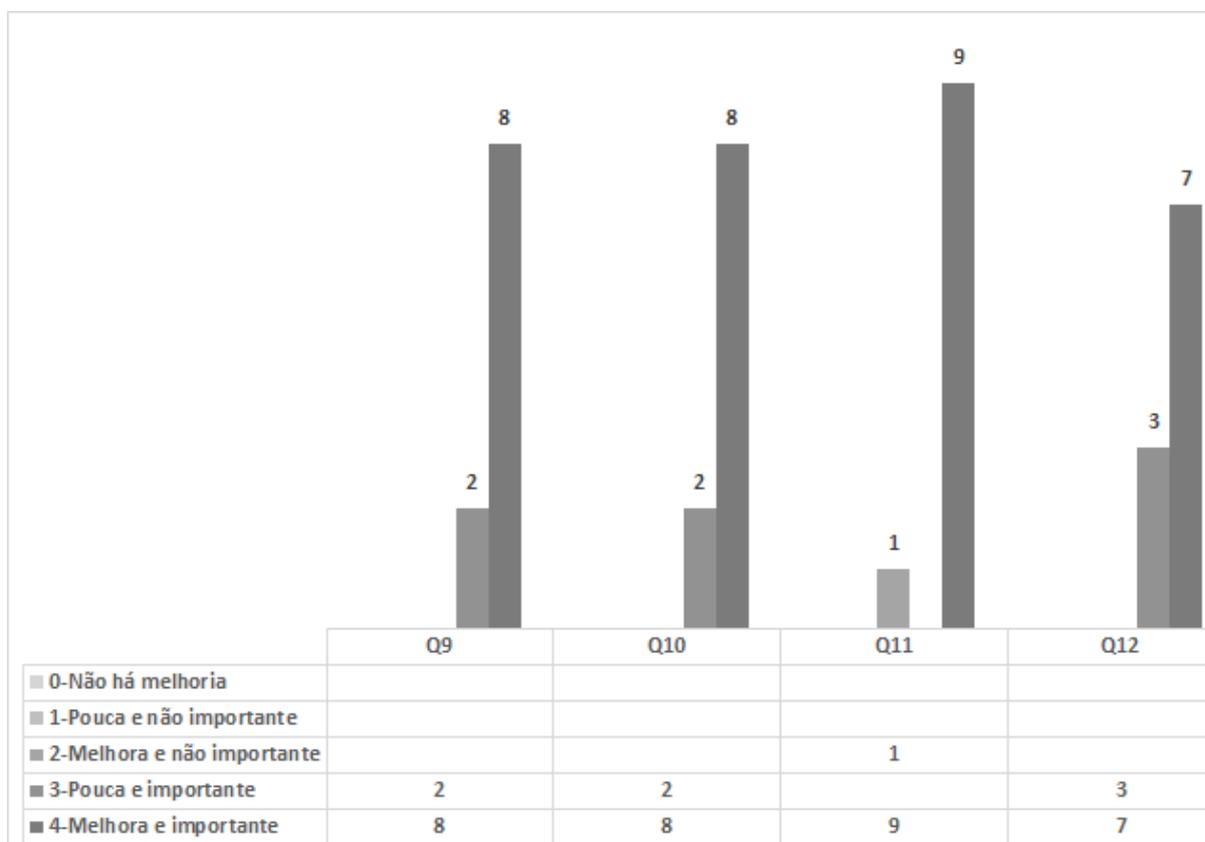


Figura 5.3: Distribuição das Manifestações por Pergunta da Segunda Seção

A pergunta Q9 indaga sobre a identificação de dados mestres e as demais perguntas da segunda seção do questionário indagam sobre o entendimento, sob diferentes perspectivas, dos domínios de dados CGU. A maioria dos especialistas (entre 7 e 9 no universo de 10) perceberam que há **BASTANTE MELHORIA** do modelo proposto em relação ao

Meta-Modelo CGU no que diz respeito a essas indagações. Além disso, quase todas as manifestações foram de que as questões postas são IMPORTANTES.

A escala de resposta para as perguntas da segunda Seção é única, permitindo aglutinar todas as manifestações às perguntas num único gráfico. Isso resume a percepção geral das melhorias do modelo para os aspectos indagados. A Figura 5.4 apresenta o resumo do número total de manifestações por cada item da escala.

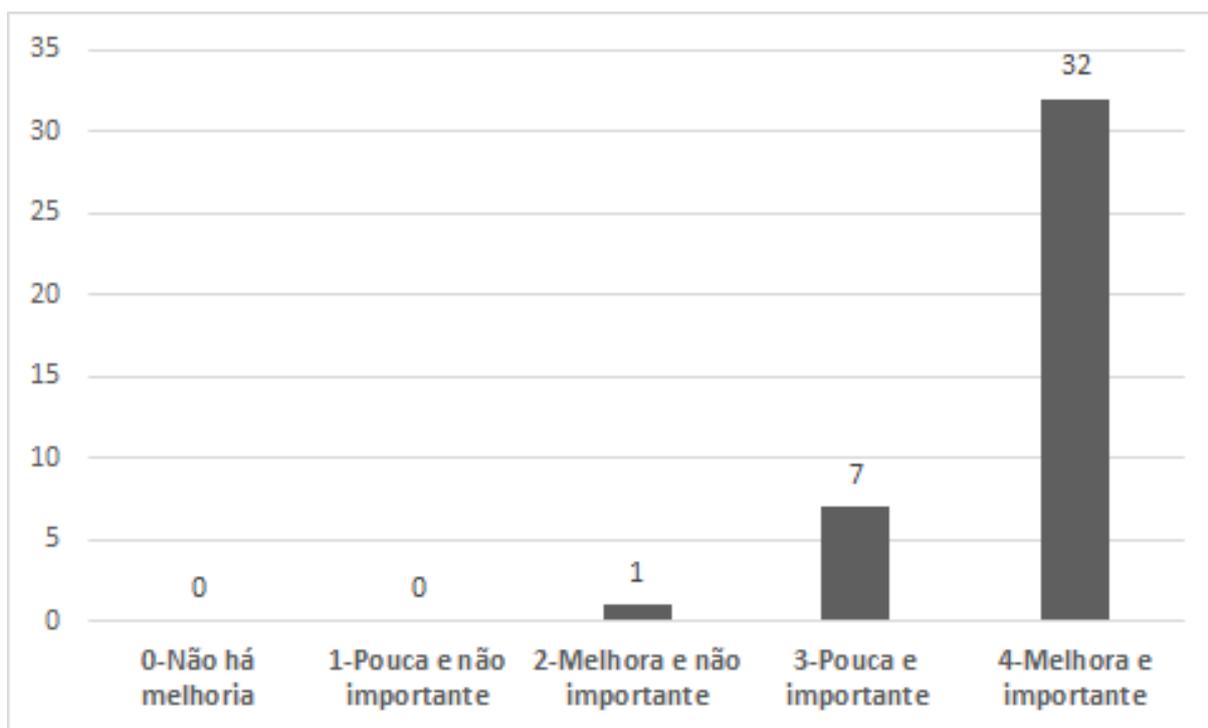


Figura 5.4: Distribuição das Manifestações às Questões da Segunda Seção

5.3 Análise

A atividade de verificação do modelo de AI por meio de mapeamentos de fontes de dados para as suas classes demonstrou que é possível mapear todas as fontes de dados para classes do modelo proposto ou estendê-lo para que comporte outras classes necessárias, utilizando a característica de extensibilidade que esse tipo de modelo deve ter.

Por outro lado, a avaliação efetuada por especialistas por meio de questionário demonstra a percepção de que o modelo baseado em ontologia é capaz de agregar melhorias importantes ao modelo estilo ER originalmente proposto (Meta-Modelo CGU), algo já esperado.

Os resultados do mapeamento de fontes e do questionário de avaliação foram considerados satisfatórios e permitem concluir que a proposta de AI deste trabalho é factível e pode

auxiliar o atendimento dos objetivos de controle PO2.1 e PO2.2 do processo Cobit PO2-Definir a Arquitetura da Informação, notadamente no que diz respeito ao entendimento e ao compartilhamento de dados.

5.4 Síntese do Capítulo

Este Capítulo apresentou os resultados de mapeamentos de fontes de dados reais da CGU para classes do modelo de AI proposto e resultados de questionário aplicado para avaliar esse modelo. Os resultados mostram que o modelo proposto é factível e é capaz de trazer melhorias ao Meta-Modelo originalmente proposto pelo Grupo de Trabalho PO2 CGU. Tendo em vista os resultados obtidos, parece válido o exame das propostas apresentadas neste trabalho no momento de implementação do processo PO2 na CGU.

Capítulo 6

Conclusão e Trabalhos Futuros

Este trabalho desenvolveu estudos sobre o tema abordagens ontológicas na gestão e integração de dados e elaborou uma proposta de arquitetura da informação (AI), sob a forma de modelo conceitual baseado em ontologia, de uma arquitetura de solução de dicionário de dados (DD) para suportar esse modelo de AI na CGU. O propósito da arquitetura de solução de DD é demonstrar o ambiente no qual o modelo de AI poderá ser implementado, servindo de base para a decisão sobre a adoção do modelo e de seus respectivos impactos.

O modelo de AI foi validado de duas formas: por meio de mapeamentos das fontes de dados para classes do protótipo do modelo, construído em utilizando uma ferramenta de ontologia, e por avaliações de especialistas da CGU. As validações permitiram concluir que o modelo de AI proposto é factível e pode auxiliar o atendimento dos objetivos de controle PO2.1 e PO2.2 do processo Cobit 4.1 – PO2 Definir a Arquitetura da Informação, particularmente no que diz respeito ao entendimento e ao compartilhamento (integração) de dados organizacionais. A percepção dos especialistas é de que ele é capaz de agregar melhorias ao Meta-Modelo anteriormente proposto para a CGU, que ainda não se encontra implementado.

A análise dos estudos existentes identificou duas propostas referenciais para o modelo de AI resultante deste trabalho: 1) a Abordagem Ontológica de Gestão de Dados (OBDM) que propõe a gestão de dados baseada em três camadas (uma de ontologia, uma de fontes de dados e outra específica para os mapeamentos entre elementos das duas camadas anteriores); 2) a arquitetura de referência holística de [2] para estrutura ontológica. Essa arquitetura holística, além de permitir uma visão abrangente e organizada dos ativos de dados de uma organização, dedica parte específica de sua estrutura para abrigar dados mestres ou corporativos (dados transversais a várias aplicações), sendo essencial para a integração de dados. As outras partes que compõem a estrutura ontológica são: uma camada de domínio de negócios, uma camada de tarefas e serviços gerais e uma camada de aplicações.

Complementarmente, este trabalho analisou algumas iniciativas relacionadas à gestão de dados e informações existentes na Administração Pública Federal, a fim de verificar a existência de modelos prontos e aptos a atender às necessidades da CGU. Apesar de as iniciativas identificadas não se encontrarem finalizadas e nem apresentarem modelos prontos e adequados para a CGU, foram identificadas algumas diretrizes e definições que foram consideradas na proposta elaborada neste trabalho.

A proposta de arquitetura da informação e a proposta de arquitetura de solução de dicionários de dados podem ser analisadas quando da implementação da governança de dados na CGU por vir, e, por sua generalidade, também podem servir de referenciais para outros órgãos com necessidades semelhantes às da CGU.

Por fim, as abordagens tipo OBDM fornecem um tópico de pesquisa interessante que tem potencial para novas atividades de pesquisa e aplicação prática nas organizações privadas ou governamentais. Os resultados podem ser utilizados, por exemplo, para a tomada de decisão em casos particulares sobre a melhor abordagem a ser adotada na gestão e governança de dados de uma organização.

6.1 Contribuições

A principal contribuição deste trabalho pode ser segmentada em:

- Fornecer um panorama das pesquisas relacionadas à utilização de ontologia para a integração e gestão de dados, bem como para a interoperabilidade entre sistemas, a fim de fundamentar a opção por abordagens ontológicas;
- A partir dos estudos selecionados na literatura, identificar os modelos baseados em ontologia que possam ser utilizados, diretamente ou como referenciais, para uma efetiva gestão de dados corporativos de governo voltada à integração;
- apresentar protótipo desenvolvido para instituição governamental com base em modelos ontológicos utilizados como referenciais;
- demonstrar a aplicabilidade desse tipo de modelo conceitual de dados por meio da validação do protótipo elaborado.

6.2 Trabalhos Futuros

Como trabalhos futuros, há interesse nas seguintes atividades:

1. aprimorar o esqueleto de AI proposto, tornando-o um modelo formal, adequadamente expresso em linguagem ontológica, (por exemplo, OWL);

2. identificar e avaliar ferramentas *open source* específicas para dicionário de dados ou de gerenciamento de dados mestres que possam compor o ambiente da arquitetura de solução de DD proposta;
3. estudar as técnicas existentes para a conversão de modelos ontológicos em modelos relacionais;
4. identificar e avaliar ferramentas *open source* que possam efetuar correspondência automática de estruturas de dados e que possam compor o ambiente de arquitetura de solução de DD proposta.

A implementação do ambiente completo para a solução de DD, mesmo que para piloto, permitirá a definição e a avaliação dos processos para a gestão de dados baseada em ontologia, bem como a verificação do suporte fornecido por essa abordagem para o entendimento e a integração de dados, e para a consequente interoperabilidade entre sistemas.

Referências

- [1] *Togaf version 9.1*. <https://www.opengroup.org/togaf/>, acesso em 2017-09-09. x, 11, 12, 13, 14, 35, 49
- [2] Fitzpatrick, Daniel, Francois Coallier e Sylvie Ratte: *A Holistic Approach for the Architecture and Design of an Ontology-Based Data Integration Capability in Product Master Data Management*, páginas 559–568. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, ISBN 978-3-642-35758-9. https://doi.org/10.1007/978-3-642-35758-9_50. x, 1, 20, 25, 26, 28, 30, 31, 48, 49, 68
- [3] CGU: *Documento de arquitetura da informação*, 2014. x, 37, 38
- [4] *Facin - modelo de referência de dados*. <http://www.participa.br/facin-framework-de-arquitetura-corporativa-e-padrees-de-interoperabilidade/visoes-do-facin/mrd-modelo-de-referencia-de-dados>, acesso em 2017-08-28. x, 40, 41, 42
- [5] Daraio, Cinzia, Maurizio Lenzerini, Claudio Leporelli, Paolo Naggari, Andrea Bonaccorsi e Alessandro Bartolucci: *The advantages of an ontology-based data management approach: openness, interoperability and data quality*. *Scientometrics*, 108(1):441–455, Jul 2016, ISSN 1588-2861. <https://doi.org/10.1007/s11192-016-1913-6>. 1, 21
- [6] Li, Yuan Fang, Gavin Kennedy, Faith Ngoran, Philip Wu e Jane Hunter: *An ontology-centric architecture for extensible scientific data management systems*. *Future Generation Computer Systems*, 29(2):641–653, 2013. 1, 20, 26, 27, 28
- [7] *Competências e organograma do ministério da transparência e controladoria-geral da união*. <http://www.cgu.gov.br/sobre/institucional/competencias-e-organograma>, acesso em 2017-08-27. 2, 50
- [8] Daraio, Cinzia, Maurizio Lenzerini, Claudio Leporelli, Henk F Moed, Paolo Naggari, Andrea Bonaccorsi e Alessandro Bartolucci: *Sapientia: the ontology of multi-dimensional research assessment*. Em *ISSI*, 2015. 2, 11, 21, 24, 26, 29, 31
- [9] *Obtain COBIT 4.1*. <http://www.isaca.org/Knowledge-Center/cobit/Pages/Downloads.aspx>, acesso em 2017-08-27. 3, 4, 11
- [10] Köhler, Jacob, Stephan Philippi e Matthias Lange: *Semeda: ontology based semantic integration of biological databases*. *Bioinformatics*, 19(18):2420–2427, 2003. 3, 6, 7, 9, 21, 24, 26, 29, 34, 56

- [11] Borst, Willem Nico: *Construction of engineering ontologies for knowledge sharing and reuse*. Universiteit Twente, 1997. 6, 47
- [12] Uschold, Michael: *Building ontologies: Towards a unified methodology*. Technical report-university of Edinburgh artificial intelligence applications institute AIAI TR, 1996. 6, 9
- [13] Guarino, Nicola: *Understanding, building and using ontologies*. International Journal of Human-Computer Studies, 46(2-3):293–310, 1997. 7
- [14] Guizzardi, Giancarlo: *Uma abordagem metodológica de desenvolvimento para e com reuso, baseada em ontologias formais de domínio*. Programa de Mestrado em Informática–Universidade Federal do Espírito Santo, Vitória, <http://wwwhome.cs.utwente.nl/~guizzard/MSc>, 2000. 7
- [15] Jones, Dean, Trevor Bench-Capon e Pepijn Visser: *Methodologies for ontology development*. 1998. 7, 8, 33
- [16] Uschold, Michael: *Converting an informal ontology into ontolingua: Some experiences*. TECHNICAL REPORT-UNIVERSITY OF EDINBURGH ARTIFICIAL INTELLIGENCE APPLICATIONS INSTITUTE AIAI TR, 1996. 8
- [17] Noy, Natalya F, Deborah L McGuinness *et al.*: *Ontology development 101: A guide to creating your first ontology*, 2001. 8
- [18] Albuquerque, Antognoni F.: *Ontological foundations for conceptual modeling datatypes*, 2013. 8
- [19] Burton-Jones, Andrew e Peter Meso: *How good are these uml diagrams? an empirical test of the wand and weber good decomposition model*. ICIS 2002 Proceedings, página 10, 2002. 8
- [20] Parsons, Jeffrey e Linda Cole: *An experimental examination of property precedence in conceptual modelling*. Em *Proceedings of the first Asian-Pacific conference on Conceptual modelling-Volume 31*, páginas 101–110. Australian Computer Society, Inc., 2004. 8
- [21] Evermann, Joerg e Yair Wand: *Toward formalizing domain modeling semantics in language syntax*. IEEE Transactions on Software Engineering, 31(1):21–37, 2005. 8
- [22] Li, Xueming e Jeffrey Parsons: *Ontological semantics for the use of uml in conceptual modeling*. Em *Tutorials, posters, panels and industrial contributions at the 26th international conference on Conceptual modeling-Volume 83*, páginas 179–184. Australian Computer Society, Inc., 2007. 8
- [23] Jasper, Robert, Mike Uschold *et al.*: *A framework for understanding and classifying ontology applications*. Em *Proceedings 12th Int. Workshop on Knowledge Acquisition, Modelling, and Management KAW*, volume 99, páginas 16–21, 1999. 9
- [24] Russo, Mariza: *Fundamentos de biblioteconomia e ciência da informação*. Número 1. Editora E-papers, 2010. 9

- [25] Ullman, Jeffrey D: *Information integration using logical views*. Em *International Conference on Database Theory*, páginas 19–40. Springer, 1997. 9
- [26] Lenzerini, Maurizio: *Data integration: A theoretical perspective*. Em *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, páginas 233–246. ACM, 2002. 10
- [27] Calvanese, Diego, Giuseppe De Giacomo, Maurizio Lenzerini, Daniele Nardi e Riccardo Rosati: *Description logic framework for information integration*. Em *KR*, páginas 2–13, 1998. 10
- [28] Loshin, David: *Master data management*. MK/OMG Press. Elsevier, Burlington, MA, 2008. <http://cds.cern.ch/record/1252160>. 10
- [29] Calvanese, Diego, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini e Riccardo Rosati: *Tractable reasoning and efficient query answering in description logics: The dl-lite family*. *Journal of Automated reasoning*, 39(3):385–429, 2007. 11
- [30] Poggi, Antonella, Domenico Lembo, Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini e Riccardo Rosati: *Linking data to ontologies*. Em *Journal on data semantics X*, páginas 133–173. Springer, 2008. 11
- [31] Lenzerini, Maurizio: *Ontology-based data management*. Em *Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM '11*, páginas 5–6, New York, NY, USA, 2011. ACM, ISBN 978-1-4503-0717-8. <http://doi.acm.org/10.1145/2063576.2063582>. 11, 21, 27, 28, 29, 34
- [32] 11, 21, 28, 29, 30, 31, 47, 53
- [33] *Cobit 5 portuguese*. <http://www.isaca.org/COBIT/Pages/COBIT-5-Portuguese.aspx>, acesso em 2017-08-29. 11
- [34] Jonkers, Henk, Marc M Lankhorst, Hugo WL ter Doest, Farhad Arbab, Hans Bosma e Roel J Wieringa: *Enterprise architecture: Management tool and blueprint for the organisation*. *Information Systems Frontiers*, 8(2):63–66, 2006. 12
- [35] *IEEE SA - 1471-2000 - IEEE Recommended Practice for Architectural Description for Software-Intensive Systems*. <https://standards.ieee.org/findstds/standard/1471-2000.html>, acesso em 2017-09-03. 12
- [36] *Archimate - linguagem de modelagem de arquitetura corporativa*. <http://www.opengroup.org/subjectareas/enterprise/archimate-overview>, acesso em 2018-06-10. 13
- [37] Petticrew, Mark e Helen Roberts: *Systematic reviews in the social sciences*. 2006. 16
- [38] Petersen, Kai, Robert Feldt, Shahid Mujtaba e Michael Mattsson: *Systematic mapping studies in software engineering*. Em *EASE*, volume 8, páginas 68–77, 2008. 16

- [39] Ivarsson, Martin e Tony Gorschek: *Technology transfer decision support in requirements engineering research: a systematic review of rej.* Requirements engineering, 14(3):155–175, 2009. 17, 18, 21, 24
- [40] Kitchenham, Barbara: *Procedures for performing systematic reviews.* Keele, UK, Keele University, 33(2004):1–26, 2004. 18, 24
- [41] Heiyanthuduwege, Sudath Rohitha e DD Karunaratne: *A learner oriented ontology to make effective learning management systems.* Em *Digital Information Management, 2006 1st International Conference on*, páginas 476–481. IEEE, 2006. 20, 26
- [42] Rui, Yang, Chen Nengcheng e Liu Zhixue: *A new approach to a local e-government portal for information management and deep searching.* Wuhan University Journal of Natural Sciences, 11(5):1161–1166, 2006. 20, 24, 25
- [43] Khan, Wajahat Ali, Asad Masood Khattak, Sungyoung Lee, Maqbool Hussain, Bilal Amin e Khalid Latif: *Achieving interoperability among healthcare standards: building semantic mappings at models level.* Em *Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication*, página 101. ACM, 2012. 20, 26
- [44] Kang, Dongwoo, Jeongsoo Lee, Sungchul Choi e Kwangsoo Kim: *An ontology-based enterprise architecture.* Expert Systems with Applications, 37(2):1456–1464, 2010. 20
- [45] Han, Yaoci e Wei Xu: *An ontology-oriented decision support system for emergency management based on information fusion.* Em *Proceedings of the 1st ACM SIGSPATIAL International Workshop on the Use of GIS in Emergency Management, EM-GIS '15*, páginas 15:1–15:8, New York, NY, USA, 2015. ACM, ISBN 978-1-4503-3970-4. <http://doi.acm.org/10.1145/2835596.2835602>. 20, 26
- [46] El Idrissi, Bouchra, Salah Baïna e Karim Baïna: *Automatic generation of ontology from data models: a practical evaluation of existing approaches.* Em *Research Challenges in Information Science (RCIS), 2013 IEEE Seventh International Conference on*, páginas 1–12. IEEE, 2013. 20, 24
- [47] Lee, Taehee, Ig hoon Lee, Suekyung Lee, Sang goo Lee, Dongkyu Kim, Jonghoon Chun, Hyunja Lee e Junho Shim: *Building an operational product ontology system.* Electronic Commerce Research and Applications, 5(1):16–28, 2006. 21, 26
- [48] Yusuf, Jawahir Che Mustapha, Patrice Boursier, Mazliham Mohd Su'ud e Muhammad Alam: *Extensive overview of an ontology-based architecture for accessing multi-format information for disaster management.* Em *Information Retrieval & Knowledge Management (CAMP), 2012 International Conference on*, páginas 294–299. IEEE, 2012. 21, 24, 25, 27
- [49] Skevakis, Giannis, Konstantinos Makris, Varvara Kalokyri, Polyxeni Arapi e Stavros Christodoulakis: *Metadata management, interoperability and linked data publishing support for natural history museums.* International Journal on Digital Libraries, 14(3-4):127–140, 2014. 21, 26

- [50] Braga, Daniele, Diego Calvanese, Alessandro Campi, Stefano Ceri, Florian Daniel, Davide Martinenghi, Paolo Merialdo e Riccardo Torlone: *Ngs: a framework for multi-domain query answering*. Em *Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on*, páginas 254–261. IEEE, 2008. 21
- [51] Park, DaeWon, Hyuk Chul Kwon e Suhyun Park: *Ontology for semantic representation of marine metadata*. Em *Proceedings of the 7th International Conference on Ubiquitous Information Management and Communication*, página 66. ACM, 2013. 21
- [52] Miah, Shah Jahan, Hafizul Islam e Ahmad ZH Samsudin: *Ontology techniques for representing the problem of discourse: Design of solution application perspective*. Em *Computer and Information Technology (CIT), 2016 IEEE International Conference on*, páginas 148–153. IEEE, 2016. 21, 24
- [53] Aminah, Siti, Iis Afriyanti e Adila Krisnadhi: *Ontology-based approach for academic evaluation system*. Em *Data Engineering (ICDE), 2017 IEEE 33rd International Conference on*, páginas 1569–1574. IEEE, 2017. 21, 24, 25, 26
- [54] Du, Juan e Vijayan Sugumaran: *Ontology-based information integration and decision making in prefabricated construction component supply chain*. 2017. 21, 24, 25, 26
- [55] Tutcher, Jonathan: *Ontology-driven data integration for railway asset monitoring applications*. Em *Big Data (Big Data), 2014 IEEE International Conference on*, páginas 85–95. IEEE, 2014. 21, 25, 26
- [56] Liu, Xin, Chungjin Hu, Jianyi Huang e Feng Liu: *Opsds: a semantic data integration and service system based on domain ontology*. Em *Data Science in Cyberspace (DSC), IEEE International Conference on*, páginas 302–306. IEEE, 2016. 21, 25, 26
- [57] Wang, Ju An e Minzhe Guo: *Ovm: An ontology for vulnerability management*. Em *Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies, CSIIRW '09*, páginas 34:1–34:4, New York, NY, USA, 2009. ACM, ISBN 978-1-60558-518-5. <http://doi.acm.org/10.1145/1558607.1558646>. 21, 26
- [58] Sadiq, Muhammad Azeem, Geoff West, David A McMeekin, Lesley Arnold e Simon Moncrieff: *Provenance ontology model for land administration spatial data supply chains*. Em *Innovations in Information Technology (IIT), 2015 11th International Conference on*, páginas 184–189. IEEE, 2015. 21, 24, 26
- [59] Xu, Zheng, Lin Mei, Yunhuai Liu, Chuanping Hu e Lan Chen: *Semantic enhanced cloud environment for surveillance data management using video structural description*. *Computing*, 98(1-2):35–54, 2016. 21, 24
- [60] Eitiveni, Imairi e Dana Indra Sensuse: *Semantic web based distributed government data center*. Em *Advanced Computer Science and Information System (ICACISIS), 2011 International Conference on*, páginas 153–158. IEEE, 2011. 21

- [61] Taweel, Adel, Stuart Speedie, Gareth Tyson, Abdel R.H. Tawil, Kevin Peterson e Brendan Delaney: *Service and model-driven dynamic integration of health data*. Em *Proceedings of the First International Workshop on Managing Interoperability and Complexity in Health Systems*, MIXHS '11, páginas 11–17, New York, NY, USA, 2011. ACM, ISBN 978-1-4503-0954-7. <http://doi.acm.org/10.1145/2064747.2064752>. 21, 26
- [62] *Protégé*. <https://protege.stanford.edu/>, acesso em 2017-09-20. 24
- [63] *Owl - semantic web standards*. <https://www.w3.org/OWL/>, acesso em 2017-09-20. 26
- [64] Smith, Barry: *Ontology (science)*. 2008. 31
- [65] Fernández-López, Mariano, Asunción Gómez-Pérez e Natalia Juristo: *Methontology: from ontological art towards ontological engineering*. 1997. 34
- [66] Uschold, Mike e Michael Gruninger: *Ontologies: Principles, methods and applications*. *The knowledge engineering review*, 11(2):93–136, 1996. 34
- [67] *The open source archimate modelling tool*. <http://https://www.archimatetool.com/>, acesso em 2018-06-03. 37
- [68] *Vocabulário controlado do governo eletrônico - vcge*. <https://www.governoeletronico.gov.br/eixos-de-atuacao/governo/interoperabilidade/vocabulario-controlado-do-governo-eletronico-vcge>, acesso em 2017-08-28. 40, 43
- [69] *Padrões de interoperabilidade de governo eletrônico*. <http://eping.governoeletronico.gov.br/>, acesso em 2017-08-28. 40
- [70] *Guia metodológico para integração de dados e processos - modelo global de dados*. <http://modeloglobaldados.serpro.gov.br/modelo-global-de-dados/guia-metodologico>, acesso em 2017-08-28. 40, 45
- [71] *Zachman International - Enterprise Architecture*. <https://www.zachman.com/>, acesso em 2017-09-13. 40
- [72] *Documentos - bem-vindo ao governo eletrônico brasileiro*. <https://www.governoeletronico.gov.br/eixos-de-atuacao/governo/interoperabilidade/governanca-e-interoperabilidade/documentos>, acesso em 2017-08-28. 40
- [73] *Modelo conceitual resumido do modelo global de dados*. <http://modeloglobaldados.serpro.gov.br/modelo-global-de-dados/modelo-conceitual>, acesso em 2017-09-11. 45
- [74] *Webprotégé*. <https://webprotege.stanford.edu/>, acesso em 2018-06-05. 48
- [75] Rajabi, Zeinab, Behrouz Minaei e Mir Ali Seyyedi: *Enterprise architecture development based on enterprise ontology*. 8:85–95, agosto 2013. 49

- [76] *Facin - modelo de conteúdo*. http://www.participa.br/articles/public/0056/4482/Modelo_de_Conteudo_v1.5.pdf, acesso em 2018-06-07. 49
- [77] *The archimate® enterprise architecture modeling language*. <http://www.opengroup.org/subjectareas/enterprise/archimate-overview>, acesso em 2018-04-26. 53
- [78] Guizzardi, Giancarlo e Gerd Wagner: *Some applications of a unified foundational ontology in business modeling*. Em *Business Systems Analysis with Ontologies*, páginas 345–367. IGI Global, 2005. 56
- [79] Benbasat, Izak, David K Goldstein e Melissa Mead: *The case research strategy in studies of information systems*. *MIS quarterly*, páginas 369–386, 1987. 58
- [80] Myers, Michael D *et al.*: *Qualitative research in information systems*. *Management Information Systems Quarterly*, 21(2):241–242, 1997. 58
- [81] Hoppen, Norberto, Liette Lapointe e Eliane Moreau: *Um guia para avaliação de artigos de pesquisas em sistemas de informação*. *Read: revista eletrônica de administração*. Porto Alegre. Edição 3, vol. 2, n. 2 (set/out 1996), documento eletrônico, 1996. 59
- [82] Appolinário, Fábio: *Dicionário de metodologia científica: um guia para a produção do conhecimento científico*. Em *Dicionário de metodologia científica: um guia para a produção do conhecimento científico*. 2007. 61
- [83] JÚNIOR, Severino Domingos da Silva e Francisco José COSTA: *Mensuração e escalas de verificação: uma análise comparativa das escalas de likert e phrase completion*. *PMKT–Revista Brasileira de Pesquisas de Marketing, Opinião e Mídia*, 15:1–16, 2014. 61
- [84] Cunha, Luísa Margarida Antunes da *et al.*: *Modelos Rasch e Escalas de Likert e Thurstone na medição de atitudes*. Tese de Doutorado, 2007. 61
- [85] Morais, Carlos: *Escalas de medida, estatística descritiva e inferência estatística*. 2005. 61
- [86] Silva, NLS da e OH da Silva: *Escalas de medidas de variáveis para diagnósticos da sustentabilidade de sistema de produção agropecuários*. *Scientia Agraria Paranaensis*, 9(2):71–84, 2010. 61