



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

UMA PROPOSTA DE GOVERNANÇA DE DADOS BASEADA EM
UM MÉTODO DE DESENVOLVIMENTO DE ARQUITETURA EMPRESARIAL

Isabel Maria Francisca dos Santos

Orientadores

Sean W. M. Siqueira
Alexandre A. Andreatta

RIO DE JANEIRO, RJ – BRASIL
SETEMBRO DE 2010

UMA PROPOSTA DE GOVERNANÇA DE DADOS BASEADA EM
UM MÉTODO DE DESENVOLVIMENTO DE ARQUITETURA EMPRESARIAL

Isabel Maria Francisca dos Santos

DISSERTAÇÃO APRESENTADA COMO REQUISITO PARCIAL PARA
OBTENÇÃO DO TÍTULO DE MESTRE PELO PROGRAMA DE PÓS-
GRADUAÇÃO EM INFORMÁTICA DA UNIVERSIDADE FEDERAL DO ESTADO
DO RIO DE JANEIRO (UNIRIO). APROVADA PELA COMISSÃO
EXAMINADORA ABAIXO ASSINADA.

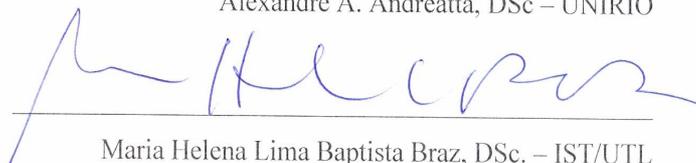
Aprovada por:



Sean Wolfgang Matsui Siqueira, DSc – UNIRIO



Alexandre A. Andreatta, DSc – UNIRIO



Maria Helena Lima Baptista Braz, DSc. – IST/UTL



Simone Bacellar Leal Ferreira, DSc – UNIRIO

RIO DE JANEIRO, RJ – BRASIL
SETEMBRO DE 2010

S237 Santos, Isabel Maria Francisca dos.
Uma proposta de governança de dados baseada em um método de desenvolvimento de arquitetura empresarial / Isabel Maria Francisca dos Santos, 2010.
x, 128f.

Orientador: Sean W. M. Siqueira.

Coorientador: Alexandre A. Andreatta.

Dissertação (Mestrado em Informática) – Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2010.

1. Sistemas de informação. 2. Governança de dados. 3. Qualidade de dados. 4. Gestão da informação. 5. TOGAF. I. Siqueira, W. M. Sean. II. Andreatta, Alexandre A. III. Universidade Federal do Estado do Rio de Janeiro (2003-). Centro de Ciências Exatas e Tecnologia Curso de Mestrado em Informática. III. Título.

CDD – 004.2

AGRADECIMENTOS

À Deus que sempre ilumina, indica o caminho certo a seguir e me carrega nos momentos mais cruciais.

Ao professor Dr. Sean W. M. Siqueira, meu orientador, pela confiança, estímulo e paciência a mim dedicada; pela competência indiscutível e por acreditar em meu potencial quando tudo parecia perdido.

Ao professor Dr. Alexandre A. Andreatta, meu co-orientador, que com dedicação e competência muito me auxiliou na conclusão deste trabalho.

Aos membros da banca examinadora, pelas contribuições para o enriquecimento desta dissertação.

Aos Professores da UNIRIO – PPGI pelos conhecimentos e atenção que me dispensaram durante o programa de mestrado e pelas sugestões e críticas durante os seminários de acompanhamento discente.

Aos funcionários da UNIRIO – PPGI que tornaram minha experiência acadêmica mais fácil e alegre.

Aos executivos e colegas de TI da empresa que trabalho, pelo apoio e contribuição.

A todas as pessoas preciosas que Deus colocou no meu caminho e que muito me incentivaram ao longo desta caminhada.

Por fim, a todos aqueles que, apesar de não serem aqui citados explicitamente, contribuíram direta ou indiretamente para a realização deste trabalho.

Santos, Isabel Maria Francisca. **Uma proposta de Governança de Dados baseada em um método de Desenvolvimento de Arquitetura Empresarial**. UNIRIO, 2010. 128 páginas. Dissertação de Mestrado. Departamento de Informática Aplicada, UNIRIO.

RESUMO

As empresas estão descobrindo que a qualidade dos dados tem um impacto significativo sobre suas iniciativas corporativas mais estratégicas e que a falta de atenção para com questões relativas à qualidade de dados é um claro inibidor de sucesso. Também estão reconhecendo que o modo como administram seus dados é crítico para sua capacidade de melhorar a agilidade e a produtividade, reduzir custos e oferecer suporte a vários padrões de armazenamento e recuperação. Entretanto poucas organizações, tanto de pequeno como de grande porte, têm uma estratégia bem definida para tratamento dos dados. Muitas vezes o dado é considerado como um bem que é criado por um departamento e devidamente guardado pelo mesmo, como se só a ele pertencesse.

Governança de dados surge como uma ação multidisciplinar que tem por objetivo tratar os dados como um recurso ativo e tangível na organização. Isto inclui políticas, padronizações, processos, pessoas e tecnologia, elementos essenciais à administração dos dados. Governança de dados também inclui uma coordenação e monitoramento para assegurar a observância ao programa e administrar riscos, bem como influir na cultura organizacional para que esta incorpore um tratamento maduro da informação, em que prioridades são definidas e os indivíduos passem a se reconhecer como gestores da informação. Este trabalho traz uma reflexão sobre o tratamento dos dados e a garantia de sua qualidade nas organizações. Também propõe um framework baseado no TOGAF-ADM, metodologia que se utiliza de uma sequência iterativa de passos para criação e desenvolvimento de uma arquitetura empresarial completa, sequência esta capaz de elucidar todos os passos necessários para uma efetiva Governança de Dados em uma organização.

Palavras-chave: Governança de Dados, Qualidade de Dados, Arquitetura Empresarial, TOGAF, Sistemas de Informação

ABSTRACT

Companies are discovering that data quality has a significant impact on their main strategic corporate initiatives, and that lack of attention to issues of data quality is a clear inhibitor of success. They are also recognizing that the way they manage their data is critical to their ability to improve agility and productivity, reduce costs and support multiple standards for storage and retrieval. However, few organizations, both small and large, have a clear strategy for data processing. Often the data is considered as an asset that is created by a department and duly recorded by the same, just as if it belongs only to it.

Data governance emerges as a multidisciplinary activity that aims to treat the data as a tangible asset and resource throughout the organization. This includes policies, standardization, processes, people and technology, the key elements for data administration. Data governance also includes coordination and monitoring to ensure compliance with the program, risk management and influencing the organizational culture so that it incorporates a mature information processing, in which priorities are set and people start to recognize themselves as information managers. This paper presents a reflection on data processing and quality assurance in organizations. It also proposes a framework based on TOGAF-ADM, a methodology that uses an iterative sequence of steps for creating and developing a complete enterprise architecture, providing an iterative sequence able to elucidate all the steps needed for effective data governance in an organization.

Keywords: Data Governance, Data Quality, Enterprise Architecture, TOGAF, Information Systems

SUMÁRIO

| | |
|--|----|
| 1. Introdução | 1 |
| 1.1 Motivação..... | 1 |
| 1.2 Problema | 2 |
| 1.3 Hipótese | 3 |
| 1.4 Objetivo do estudo | 3 |
| 1.5 Metodologia | 3 |
| 1.6 Limitações da pesquisa | 4 |
| 1.7 Organização do trabalho | 4 |
| 2. Conceitos Fundamentais | 5 |
| 2.1 Ambiente de Apoio à Decisão..... | 5 |
| 2.1.1 Conceitos e características | 6 |
| 2.1.2 Credibilidade dos dados | 7 |
| 2.2 Qualidade de Dados e Informação | 8 |
| 2.2.1 Visão geral | 8 |
| 2.2.2 Critérios de qualidade de dados | 10 |
| 2.2.3 Metodologias relacionadas à qualidade de dados | 14 |
| 2.3 Governança de Dados | 19 |
| 2.3.1 Programa de Governança de Dados proposto pela Oracle | 21 |
| 2.3.2 Programa de Governança de Dados proposta pela IBM | 23 |
| 2.3.3 Programa de Governança de Dados segundo o Data Governance Institute | 25 |
| 2.4 Arquitetura Empresarial | 27 |
| 2.4.1 Framework Zachman | 28 |
| 2.4.2 Framework TOGAF | 31 |
| 2.5 Considerações sobre Arquitetura Corporativa | 34 |
| 3. Pesquisa Exploratória: Necessidade de apoiar equipe de TI na explicitação da qualidade dos dados em DW e/ou DM (Panorama Brasil)..... | 36 |
| 3.1 Considerações Metodológicas..... | 37 |
| 3.2 Contexto da pesquisa | 38 |
| 3.3 Questionário e forma de aplicação | 39 |
| 3.4 Perfil dos Respondentes | 40 |
| 3.5 Resultados obtidos | 41 |
| 3.6 Conclusão e comentários finais..... | 51 |
| 4. DATAGOV-ADM: Uma proposta de Governança de Dados baseada no TOGAF-ADM52 | |
| 4.1 Iniciativas estratégicas | 53 |
| 4.1.1 Definição de papéis e responsabilidades..... | 56 |
| 4.1.2 Definição de políticas e procedimentos | 57 |
| 4.2 Visão e Objetivos | 58 |
| 4.2.1 Definições e Fundamentação estratégica | 58 |
| 4.2.2 Regulamentações e <i>Compliance</i> | 59 |
| 4.2.3 Gestão de Risco Operacional e a Tecnologia da informação. | 61 |
| 4.3 Processos de Negócio..... | 61 |
| 4.4 Sistemas de Informação | 62 |
| 4.5 Tecnologia e padronização..... | 64 |
| 4.5.1 Qualidade de Dados | 64 |
| 4.5.2 Sincronização de Dados | 70 |
| 4.5.3 Segurança dos dados | 70 |
| 4.5.4 Padronização de Dados | 72 |

| | |
|---|-----|
| 4.6 Oportunidades e soluções..... | 73 |
| 4.7 Plano de implementação | 73 |
| 4.7.1 Cultura e comportamento em relação à informação..... | 74 |
| 4.7.2 Treinamento | 74 |
| 4.8 Implementação da governança..... | 75 |
| 4.9 Administração da governança..... | 76 |
| 4.10 Administração de requisitos..... | 77 |
| 5. O Estudo de Caso..... | 80 |
| 5.1 Caracterização da Empresa | 80 |
| 5.2 Caso da Empresa X..... | 81 |
| 5.3 Utilizando o framework proposto na Empresa X..... | 81 |
| 5.4 Avaliação do Método..... | 85 |
| 5.4.1 Perfil dos Respondentes | 85 |
| 5.4.2 Resultados obtidos | 86 |
| 6. Conclusão..... | 95 |
| 6.1 Comparação com outras propostas de Governança de Dados..... | 96 |
| 6.2 Contribuições da Pesquisa..... | 98 |
| 6.3 Perspectivas Futuras..... | 98 |
| Referências..... | 100 |
| ANEXO I | 107 |
| ANEXO II..... | 112 |
| ANEXO III..... | 116 |
| ANEXO IV | 120 |

LISTA DE FIGURAS

| | |
|---|----|
| Figura 1 - Fases do AIMQ | 17 |
| Figura 2 - Governança Dados da Oracle | 21 |
| Figura 3 - Elementos de uma efetiva Governança Dados segundo a IBM | 25 |
| Figura 4 - Zachman Framework | 31 |
| Figura 5 - Arquitetura Corporativa do TOGAF | 32 |
| Figura 6 - Método de desenvolvimento da arquitetura (ADM) do TOGAF | 33 |
| Figura 7 - Tempo de atuação em TI e em projetos de DW e/ou DM | 40 |
| Figura 8 - Tempo de início projeto DW e/ou DM e disponibilização primeira aplicação OLAP | 41 |
| Figura 9 - Identificam pessoas na administração que patrocinam projetos de melhoria contínua qualidade dados | 42 |
| Figura 10 - Identificam pessoas responsáveis (coordenador/executor) por ações de melhoria na qualidade dados transacionais | 42 |
| Figura 11 - Identificam pessoas pertencentes às áreas de negócio da organização envolvidas com ações de melhoria na qualidade dados | 43 |
| Figura 12 - Situação da organização ao utilizar dados com qualidade | 44 |
| Figura 13 - Missão e objetivos do DW e/ou DM encontram-se bem definidos | 44 |
| Figura 14 - Documentação das fontes de dados do projeto de DW e/ou DM | 45 |
| Figura 15 - Existência de um repositório central de metadados | 46 |
| Figura 16 - Metodologia utilizada para definir conceitos e procedimentos para a definição, medição, análise e melhoria da qualidade dos dados | 47 |
| Figura 17 - Tratamento dos atributos no processo de ETL | 47 |
| Figura 18 - Tratamento de regras de negocio no processo de ETL | 48 |
| Figura 19 - Avaliação tratamento de dados efetuado no processo de ETL | 48 |
| Figura 20 - Avaliação tratamento de dados na sumarização | 49 |
| Figura 21 - Avaliação da garantia da qualidade de dados no DW e/ou Data Mart executada pelo processo de tratamento e limpeza dos dados | 50 |
| Figura 22 - Quando um usuário manifesta necessidade de mais informações sobre a qualidade do conteúdo | 50 |
| Figura 23 – DATAGOV-ADM: Governança Dados baseada no TOGAF-ADM | 53 |
| Figura 24 - Fases na produção e utilização da informação | 55 |
| Figura 25 - Área de Atuação dos respondentes e tempo na organização | 86 |
| Figura 26 - Influência na organização ao utilizar um Framework Corporativo de Governança de Dados | 87 |
| Figura 27 - Ordem de importância das Etapas de Governança de Dados | 88 |
| Figura 28 - Características relevantes da Governança de Dados baseada em um Framework Corporativo | 90 |

| | |
|---|-----|
| Figura 29 - Características das abordagens para Qualidade de Dados, Governança de Dados e ambas | 91 |
| Figura 30 - Ciclo de vida global da organização | 123 |
| Figura 31 - Resumo arquitetura de processos corporativos | 123 |
| Figura 32 - Composição da Arquitetura FEAF | 125 |
| Figura 33 - Framework TEAF | 127 |

LISTA DE TABELAS

| | |
|--|-----|
| Tabela 1 - Agrupamento de Critérios da Qualidade de Dados | 11 |
| Tabela 2 - Quadro de Referência das propostas de critérios para a qualidade de dados | 13 |
| Tabela 3 - Metodologias para qualidade de dados | 15 |
| Tabela 4 - PSP/IQ Model | 16 |
| Tabela 5 - Avaliação da qualidade da informação no nível de atributo | 67 |
| Tabela 6 - Avaliação da qualidade da informação no nível de registros | 68 |
| Tabela 7 - Avaliação da qualidade da informação no nível de arquivo | 70 |
| Tabela 8 - Matriz da Arquitetura FEAF | 125 |

ACRÔNIMOS

| | |
|-----------|---|
| AIMQ | A methodology for Information Quality assessment |
| AMEQ | Activity-based Measuring and Evaluating of product information Quality (AMEQ) methodology |
| BI | Business Intelligence |
| CDQ | Comprehensive methodology for Data Quality Management |
| CIHI | Canadian Institute for Health Information Methodology |
| COLDQ | Cost-effect of Low Data Quality |
| CRM | Customer Resource Management |
| DaQuinCIS | Methodology for the Quatily Assessment of Financial Data |
| DGI | Data Governance Institute |
| DQA | Data Quality Assessment |
| DM | Data Mart |
| DRI | Disaster Recovery Institute |
| DW | Data Warehouse |
| DWQ | Data Warehouse Quality |
| EIS | Sistemas de informações executivas |
| ETL | Extração e Transformação e Carga |
| FTP | File Transfer Protocol |
| GQM | Goal Question Metric |
| IP | Internet Protocol |
| IQM | Information Quality Measurement |
| ISTAT | ISTAT Methodology |
| MDM | Master Data Management |
| PMI | Project Management Institute |
| SGBD | Sistema de Gestão de Base de Dados |
| SIG | Sistemas de Informação Gerenciais |
| TDQM | Total Data Quality Management |
| TI | Tecnologia da Informação |
| TIQM | Total Information Quality Management |
| TOGAF | The Open Group Architecture Framework |
| TQM | Total Quality Management |

1. Introdução

O objetivo desse capítulo é estabelecer o contexto de nossa pesquisa. Em primeiro lugar, são apresentados os aspectos que a motivam, bem como o problema e hipótese desta dissertação. Em segundo, é detalhado o objetivo que se espera atingir. Em seguida, são discutidos os aspectos de contribuição do nosso trabalho e suas limitações, no contexto de Governança e Qualidade de Dados. Finalmente, é apresentado um breve resumo do que será apresentado nos capítulos subsequentes.

1.1 Motivação

Dispor da informação correta no menor tempo possível é o grande diferencial para as empresas que querem manter-se na dianteira no mundo dos negócios. É o que lhes permite tomar decisões rápidas, corrigir rumos, adequar-se às oscilações econômicas e a questões legais, e antecipar-se às necessidades de seus clientes. Grande parte do setor corporativo percebe a necessidade de projetos de *Business Intelligence* (BI). Entretanto, ainda é muito alto o índice de projetos que não obtém êxito ou não atendem as expectativas e necessidades da organização. As estatísticas revelam que mais da metade dos projetos de BI não são concluídos, ou fracassam, consumindo milhões de dólares sem trazer os resultados esperados. Isso acontece por uma sucessão de erros, começando pelo desconhecimento do que de fato é *Business Intelligence* (FRIEDMAN, et.al., 2009). Grande parte das empresas ainda considera projetos de BI como projetos de Tecnologia da Informação (TI); não percebem BI como um conceito atrelado à estratégia corporativa, e que tem como principal foco transformar os dados (coletados pelos sistemas transacionais) em informações que subsidiem a tomada de decisões.

O cuidado com o tratamento dos dados é um dos elementos fundamentais para que um projeto de BI não fracasse. São poucas as corporações que adotam uma prática estruturada de gestão de dados e, sem essa prática, há grande risco de ocorrerem falhas nas iniciativas de BI. A baixa qualidade dos dados resulta em relatórios imprecisos, análises incorretas de posicionamento no mercado, pedidos reprocessados, ineficiência das áreas de marketing e comercial, custos superiores à média do mercado, duplicação de informações, dependência excessiva do conhecimento tácito dos tomadores de decisão, entre outros. Em especial dados incorretos acarretam vários problemas como, por exemplo: repetição de trabalho para efetuar correções; esforço exagerado para elaboração de rotinas de extração de dados; perda de oportunidades de negócios e clientes; risco à imagem da organização e até mesmo multas por não cumprimento de normas regulatórias.

Projetos de BI incorporam iniciativas que visam a melhoria da qualidade dos dados nas organizações e buscam implementar algum processo de melhoria contínua da qualidade de seus dados. Projetos de BI e qualidade de dados devem ser tratados pelas organizações dentro de um programa de Governança de Dados, de ação contínua e que envolva a alta direção, os gestores das áreas de negócio, e os profissionais da TI envolvidos com administração e qualidade dos dados. Um programa de Governança de Dados é um conjunto de processos que assegura que ativos de dados importantes sejam gerenciados formalmente através da empresa. A Governança de Dados assegura que os dados sejam confiáveis, e que pessoas possam ser responsabilizadas por qualquer evento adverso que ocorrer por conta de baixa qualidade dos dados. Fisher (2009) afirma que o problema relacionado dos dados é um problema de todos os executivos e empregados em uma organização. A Governança de Dados seria muito mais do que um programa, seria sim um “estilo de vida” da organização. O autor acrescenta que toda organização não deveria medir esforços para implementar a Governança de Dados de forma a adquirir maior competitividade para vencer os desafios atuais.

1.2 Problema

As empresas vêm reconhecendo que o modo como administram seus dados é crítico para sua capacidade de melhorar a agilidade e a produtividade, reduzir custos e oferecer suporte a vários padrões de armazenamento e recuperação. A ausência de um programa de Governança de Dados tem um impacto significativo sobre suas iniciativas

corporativas mais estratégicas e a falta de atenção para com questões relativas à qualidade de dados é um claro inibidor de sucesso. Logo, é necessário que as organizações adotem um programa de Governança de Dados que alinhe TI e negócio, e que utilizem alguma metodologia para orientar esta questão.

1.3 Hipótese

A Arquitetura Empresarial, por ser um grande modelo que documenta e representa todos os comportamentos que fazem uma organização funcionar de forma integrada e sincronizada, pode servir de base para uma metodologia de Governança de Dados que alinhe a tecnologia da informação e o negócio, utilizando uma sequência iterativa capaz de elucidar todos os passos necessários para sua implementação.

1.4 Objetivo do estudo

O objetivo desse estudo é apresentar uma proposta de um “*framework*” de Governança de Dados baseado no TOGAF (2009), que é uma metodologia que emprega uma sequência iterativa de passos para a criação e desenvolvimento de uma arquitetura empresarial completa. Tal *framework* utiliza um conjunto similar de etapas ao descrito no TOGAF, mas adaptadas ao contexto específico de Governança de Dados.

1.5 Metodologia

Foi feita uma revisão bibliográfica sobre arquitetura empresarial, Governança de Dados e iniciou-se a confecção de um framework de Governança de Dados capaz de permitir que a informação seja tratada como um recurso tangível, que deva ser reconhecido e tratado como qualquer outro recurso tangível existente na organização. Em seguida, foi feito um estudo de caso de utilização deste framework em uma empresa do ramo de previdência complementar, para corroborar a possibilidade de implementação e avaliar os resultados, mesmo que em um escopo de trabalho bem reduzido. Dentro deste escopo, foram executadas todas as fases propostas neste método. Ao final da execução de uma iteração da proposta de *framework*, o resultado obtido foi apresentado a um grupo de 9 pessoas: 4 Analistas de Sistemas, 1 Analista de Segurança, 1 Administrador de Dados, 1 Economista, 2 Administradores de Empresas, permitindo assim que fosse

feita uma avaliação do método proposto tanto no âmbito da TI como por profissionais da área de negócio (Anexo II).

1.6 Limitações da pesquisa

Houve dificuldade em se avaliar a aplicabilidade do framework proposto em uma organização na sua totalidade e em diferentes cenários. Assim, a pesquisa foi aplicada em um contexto específico. Além disso, também observa-se uma limitação em relação à quantidade de pessoas envolvidas na avaliação deste estudo de caso.

1.7 Organização do trabalho

Esta dissertação possui mais cinco capítulos, além deste introdutório. No segundo capítulo, apresentam-se os principais conceitos necessários ao bom entendimento do estudo desenvolvido e da metodologia proposta. No terceiro capítulo, apresenta-se a pesquisa exploratória desenvolvida quanto à necessidade de apoiar as equipes de TI na explicitação da qualidade dos dados em projetos de construção de DW e/ou DM. Tal pesquisa foi utilizada para evidenciar a necessidade da inserção de tais projetos dentro de um programa de Governança de Dados. No quarto capítulo, especifica-se a proposta de Governança de Dados baseada no TOGAF-ADM. No quinto capítulo mostra-se, através de um estudo de caso, a aplicação da proposta descrita no capítulo anterior em uma empresa. No mesmo capítulo, descreve-se a avaliação do estudo de caso descrito anteriormente por profissionais da empresa submetida ao estudo de caso, oriundos da área de TI e da área de negócios da empresa. No último capítulo, apresenta-se a conclusão do estudo, fazendo uma comparação com propostas de Governança de Dados encontradas na literatura e sugerindo a indicação de trabalhos futuros.

2. Conceitos Fundamentais

Este capítulo tem a finalidade de apresentar os conceitos necessários para o entendimento deste trabalho. Deste modo, inicia-se discutindo conceitos e características de um projeto de apoio à decisão e a importância da credibilidade dos dados apresentados para o sucesso desse projeto. Dando prosseguimento e como qualidade de dados e informação está profundamente ligada à Governança de Dados, os principais critérios de qualidade de dados são discutidos bem como as metodologias relacionadas à qualidade de dados encontrada na literatura.

O conceito de Governança de Dados, é definido, bem como alguns fatores críticos para o sucesso desta governança. Também se encontram descritos alguns trabalhos que propõem programas de Governança de Dados. E por último, como o alinhamento proposto entre o negócio e a área de Tecnologia da Informação é baseado em arquiteturas empresariais, faz-se uma breve colocação sobre a importância e os benefícios de utilizá-las e detalha-se o framework TOGAF do *The Open Group Architecture* (THE OPEN GROUP, 2008) que serviu como base para a proposta desenvolvida neste trabalho.

2.1 Ambiente de Apoio à Decisão

Informações importantes ficam armazenadas dentro da própria empresa, sendo seu acesso muito difícil, devido à existência de múltiplos estoques de dados, com lógicas próprias e dificuldades de consolidação. O conceito de *Business Intelligence* (BI) está relacionado com a resolução conjunta destes problemas, através de um ambiente de conhecimento, onde haja produção sistemática de informação gerencial, veloz e consistente. Tal informação deve ser capaz de auxiliar aos gestores das empresas a tomarem as melhores decisões nos negócios, baseado em fatos reais, com profundidade suficiente para descobrir as causas de uma tendência ou de um problema. A seguir

detalha-se o conceito e as características deste ambiente e como a credibilidade dos dados utilizados é fundamental para seu sucesso.

2.1.1 Conceitos e características

BI - *Business Intelligence* é um termo genérico que inclui arquiteturas, ferramentas, bancos de dados, aplicações e metodologias (RAISINGHANI, 2004). Este termo foi cunhado pelo *Gartner Group* em meados da década de 90. Contudo, o conceito teve início muito antes, com suas raízes nos sistemas de geração de relatórios (Sistemas de Informação Gerenciais – SIG) dos anos 70. Durante este período, os sistemas de geração de relatórios eram baseados em planilhas estáticas e não possuíam recursos de análise. No início dos anos 80, surgiu o conceito de sistemas de informações executivas (EIS) que traziam novos recursos tais como: geração de relatórios e planilhas dinâmicas (*ad hoc* ou sob demanda), prognósticos e previsões, análise de tendências, detalhamento, acesso a status e fatores críticos de sucesso. Esses recursos apareceram em diversos produtos comerciais até meados da década de 90. Mais tarde os mesmos recursos e alguns novos apareceram sob o nome de BI.

Os principais objetivos do BI são permitir o acesso interativo aos dados, às vezes em tempo real, proporcionar a manipulação desses dados e fornecer aos gerentes e analistas de negócios a capacidade de realizar a análise adequada. Ao analisarem dados, situações e desempenhos históricos e atuais, os tomadores de decisão conseguem valiosos *insights* que podem servir como base para decisões melhores (ZAMAN, 2005).

Data Warehouse (INMON, 2003) é uma coleção de dados atuais e históricos, orientada por assunto, integrada, variável no tempo e não-volátil, de interesse aos gerentes de toda a organização, possibilitando assim oferecer suporte à tomada de decisões. Os dados normalmente são estruturados de modo a estar disponíveis em um formato pronto para as atividades de processamento analítico on-line (OLAP), mineração de dados, consultas, geração de relatórios e demais aplicações de suporte à decisão.

O *Data Mart* (KIMBALL, 2000) é uma base de dados que contém dados operacionais específicos de uma determinada área ou assunto, e que apoiam a tomada de decisão de um grupo particular de usuários. Logo, é comum que uma empresa possua DMs específicos para determinados departamentos e setores.

Utilizando-se qualquer uma das abordagens anteriores, persiste a demanda por qualidade e credibilidade dos dados que estão sendo coletados, conforme será abordado na próxima seção.

2.1.2 Credibilidade dos dados

A credibilidade dos dados é muito importante para o sucesso de um projeto. Discrepâncias simples podem causar sérios problemas quando se deseja extrair dados para suportar decisões estratégicas para o negócio das empresas. Dados não dignos de confiança podem resultar em relatórios inúteis, como por exemplo uma lista de pacientes do sexo masculino e grávidos. Se existirem dados de má qualidade e os mesmos forem disponibilizados em um DW, o resultado final será um suporte à decisão de baixo nível com altos riscos para o negócio. Coisas aparentemente simples, como um Código de Endereçamento Postal (CEP) errado, podem não ter nenhum impacto em uma transação de compra e venda, mas podem influir nas informações referentes à cobertura geográfica, por exemplo.

Não é apenas a escolha da ferramenta certa que influi na qualidade dos dados. Segundo (DATA WAREHOUSING INSTITUTE, 2006), conjuntos de coleções de dados, processos de entrada, metadados e informações sobre a origem dos dados, são importantíssimos. Outras questões como a manutenção e atualização dos dados e as diferenças entre dados para bancos transacionais e para uso em Data Warehousing também são cruciais para o sucesso dos projetos.

A próxima seção descreve as motivações para se preservar a qualidade dos dados da organização, assim como faz uma revisão e atualização dos critérios de qualidade de dados definidos por diversos autores, e enumera as metodologias de explicitação dessa qualidade.

2.2 Qualidade de Dados e Informação

2.2.1 Visão geral

Pode-se verificar que a qualidade é um conceito complexo, porque possui significados diversos para diferentes pessoas, em um contexto altamente dependente. Para a norma brasileira NBR ISO 8402 (1994), considera-se qualidade: “A totalidade das características de uma entidade que lhe confere a capacidade de satisfazer às necessidades explícitas e implícitas”.

De acordo com o dicionário Aurélio (FERREIRA, A.B.H, 1999), Qualidade é “Propriedade, atributo ou condição das coisas ou das pessoas capaz de distingui-las das outras e de lhes determinar a natureza”. Ainda no dicionário Aurélio, pode-se encontrar outra definição: “Numa escala de valores, qualidade permite avaliar e, conseqüentemente, aprovar, aceitar ou recusar, qualquer coisa”. Já a ISO 9000 (2005) a define como “... o grau no qual um conjunto de características inerentes satisfaz a requisitos”.

Segundo BELCHIOR (1992), *“Qualidade é uma característica intrínseca e multifacetada de um produto. A relevância de cada faceta pode variar com o contexto e ao longo do tempo, pois as pessoas podem mudar seus posicionamentos e atualizar seus referenciais, com relação a um objeto ou a uma questão. Portanto, a qualidade não é absoluta, mas depende da perspectiva do avaliador”*.

Qualidade de dados é um conceito complexo, com significados diversos para diferentes pessoas. Analisando-se estudos já realizados e propostas para a definição de critérios para a qualidade de dados, pode-se observar essa diversidade, não havendo ainda um consenso em relação a um conjunto de critérios que pudesse ser sempre utilizado para definir qualidade. Uma razão para isso é o caráter essencialmente subjetivo da qualidade de dados, cuja avaliação pode variar de acordo com a função do observador, do contexto e dos objetivos da avaliação. Além disso, muitas vezes, a qualidade não pode ser expressa quantitativamente, mas precisa ser descrita (AMARAL, 2003).

Para analisar como uma organização adota práticas de gestão de dados, há uma série de perguntas a serem elaboradas. Uma delas é se a empresa tem problemas, custos adicionais ou perdas financeiras, ficando exposta a riscos em decorrência da baixa qualidade de seus dados. Outra, por exemplo, é o nível de dependência que a empresa

tem de processos automatizados de tomada de decisão. São também considerados indicativos do comprometimento com gerenciamento de dados a atenção que a alta administração dá ao tratamento de dados e a existência de uma estratégia para o assunto (CAMPOS, 2005).

Para que a função de administração de dados seja efetiva, padrões e políticas a respeito dos dados, sua definição e seu uso devem ser adotados. Esses padrões devem ser rigorosos, abrangentes e flexíveis a mudanças, para poder viabilizar a reutilização, a estabilidade e a efetiva comunicação do significado dos dados. Deve-se usar ferramentas, como dicionário de dados e repositórios para gerenciá-los (XAVIER, et.al., 1999).

Dados bem definidos, íntegros no seu significado, consistentes, confiáveis, seguros e compartilhados fazem com que cada novo sistema defina apenas os dados que são do seu escopo e divida os demais com outros sistemas presentes na organização (XAVIER, et.al., 1999). E o que é arriscado para as organizações é que a qualidade dos dados tende a piorar e, deste modo, é preciso uma atenção especial, por meio de métodos, técnicas e ferramentas específicas (FILHO, 2003).

O conceito de qualidade de dados indica o quanto a fontes de dados atendem às demandas de informação das áreas de negócio. Isto envolve conhecer a qualidade atual de dados, determinar a qualidade desejada de dados e ajustar a qualidade atual em função da desejada (FILHO, 2003).

Logo, gerar uma medida de qualidade aceitável para todos não é trivial. Para isso é preciso definir as características da qualidade que interessam e, então, decidir como a avaliação da qualidade será feita através delas (PINHO, 2001).

Um sistema de informação de qualidade pressupõe um sistema que cumpre seus objetivos, é gerenciável, é passível de manutenção e de aprendizado por uma pessoa que não tenha feito parte do grupo original do projeto (CAMPOS, 2005). À medida que a infra-estrutura de informações das empresas amadurece, aumenta a necessidade de qualidade das próprias informações e de sistemas eficientes e eficazes de suporte à decisão, os quais são projetados para apoiar os gestores de negócio no processo de tomada de decisão numa perspectiva de mais longo prazo, no trato da informação (FILHO, 2003).

2.2.2 Critérios de qualidade de dados

O conceito de qualidade de dados, conforme apresentado na literatura, é um conceito multidimensional. Acurácia, completeza, consistência, confiabilidade, veracidade, relevância e atualidade são critérios freqüentemente citados. Contudo, ainda não existe um consenso em relação a esses critérios.

Uma abordagem para a definição de critérios da qualidade, usando conceitos ontológicos, foi utilizada por WAND e WANG (1996). Essa abordagem baseia-se nos problemas que acontecem no mapeamento dos dados do mundo real para os sistemas de informação. Seu principal objetivo é fornecer orientações para os projetistas de sistemas, em questões relacionadas à qualidade de dados. O resultado do trabalho realizado por WAND e WANG (1996), contém os critérios para a qualidade de dados nos sistemas de informação, relacionados às visões interna (ou seja, do ponto de vista técnico) e externa (isto é, do ponto de vista dos usuários).

Ainda segundo estes pesquisadores, na visão interna, a qualidade de dados está diretamente ligada à confiabilidade, temporalidade, completeza, atualidade, consistência e precisão; sob o ponto de vista dos usuários (visão externa), a qualidade dos dados está diretamente ligada a aspectos de temporalidade, relevância, conteúdo, importância, suficiência, facilidade de uso, utilidade, clareza, concisão, ser informativo, nível de detalhe, quantidade apropriada, escopo, facilidade de interpretação e entendimento.

Segundo ECKERSON (2002), o dado não precisa estar totalmente livre de defeitos. Basta que ele atenda aos requisitos dos usuários. Por este motivo, os critérios que tem como objetivo caracterizar a qualidade dos dados, de uma forma geral, são: acurácia, integridade, consistência, completeza, validade dos domínios, disponibilidade e acessibilidade.

O objetivo principal da pesquisa realizada por WANG et al. (1996), foi elaborar um estudo empírico, voltado para a qualificação dos dados, que capturasse aspectos de qualidade importantes para os consumidores de dados e que culminasse com a elaboração de uma arquitetura hierárquica da qualidade de dados. Como resultado, esse trabalho propôs uma tabela (Tabela 1) contendo os critérios agrupados em quatro categorias, na definição de dados de alta qualidade: intrínseca, de acessibilidade, de contexto, e aspectos representacionais.

| Categoria | Critérios |
|------------------|---|
| Intrínseca | Acurácia, objetividade, confiabilidade e reputação |
| Acessibilidade | Acessibilidade e segurança |
| Contextual | Relevância, valor agregado, “temporalidade”, completeza e quantidade apropriada de dados |
| Representacional | Facilidade de interpretação, facilidade de entendimento, consistência na representação e concisão |

Tabela 1 - Agrupamento de Critérios da Qualidade de Dados (WANG et al., 1996)

LOSHIN (2001) analisou a qualidade de dados sob o ponto de vista da gestão do conhecimento. Esse autor define qualidade de dados como "adequação para o uso" e destaca que a avaliação da qualidade é dependente do contexto do usuário. LOSHIN (2001) baseou-se em WANG et al. (1996) para definir critérios para a qualidade de dados, que foram agrupados nas seguintes categorias: modelo de dados, valores dos dados, domínios da informação, apresentação dos dados, e política da informação. Os critérios definidos por LOSHIN (2001), associados a cada uma das categorias citadas, são:

- **Modelo de Dados** - clareza na definição, facilidade de compreensão, flexibilidade, robustez, conter apenas o essencial, granularidade dos atributos, precisão dos domínios, homogeneidade, naturalidade, capacidade de identificação, facilidade de obtenção, relevância, simplicidade, consistência semântica e consistência estrutural.
- **Valores dos Dados** - acurácia, tratamento de valores nulos, completeza, consistência e atualidade;
- **Domínios da Informação** - concordância dos usuários em relação à utilização, definição de responsabilidade sobre a informação;
- **Apresentação dos Dados** - adequação aos usuários, interpretação correta, flexibilidade, precisão do formato, portabilidade, consistência da representação, representação de valores nulos e espaço de armazenamento;
- **Política da Informação** - acessibilidade, definição de metadados, privacidade, redundância, segurança e custo unitário.

Um sumário das pesquisas realizadas no mercado e na área acadêmica, sobre a definição de critérios da qualidade de dados foi desenvolvido por Lee et al. (2001) e complementado por AMARAL (2003), que atualizou o quadro de referência que

relaciona os critérios da qualidade aos autores que os propuseram. Dentre eles estão alguns autores já citados e outros que também desenvolveram pesquisas sobre a definição de critérios da qualidade de dados. A este sumário foram acrescentados os critérios propostos por WIXON e WATSON (2001), KAHN, STRONG e WANG (2002), JENNEX e OLFMAN (2002), OSLEN (2003), SHANKARANARA Yan, ZIAD e WAND (2003), WIXON e TODD (2005), KIM, KISHORE e SANDERS (2005). Esse quadro de referência revisto é apresentado na Tabela 2.

| Critério | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | Total |
|-----------------------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-------|
| Completeza | | X | X | | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | 21 |
| Temporalidade | X | X | | | X | X | X | | X | X | X | X | | X | X | X | | X | X | X | X | | X | X | X | 19 |
| Acurácia | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | | | X | | X | | | | 18 |
| Consistência | | | X | | X | | X | | X | | | X | | X | X | X | X | X | | X | X | X | | X | | 14 |
| Precisão | X | X | X | | X | | | | | | | | | | | X | X | X | X | | | X | X | X | X | 12 |
| Relevância | | X | X | | X | X | | | X | X | | | | | | X | | X | X | | X | X | | | X | 12 |
| Confiabilidade | X | X | | X | X | X | | | X | X | X | X | | | | | | | | | | X | | | | 10 |
| Facilidade de entendimento | X | X | X | | X | X | | | X | X | | | | | | X | | X | | | X | | | | | 10 |
| Acessibilidade | | | | X | | X | | X | X | X | | | X | | | | | X | | X | X | | | | X | 10 |
| Atualidade | | X | X | X | X | | | | | | | | X | | X | X | | | | | | | | | | 7 |
| Quantidade apropriada | X | X | | | X | X | | | | X | | | | | | | | X | | | X | | | | | 7 |
| Existência de metadados | | | X | X | | | X | | X | | | | | | | X | | | | | X | | | | | 6 |
| Importância | | X | X | | X | | | | | | | | | | X | X | | | | | | X | | | | 6 |
| Representação consistente | | X | X | | X | | | | X | | | | | | | X | | | | | | | | | | 5 |
| Usabilidade | | X | | X | X | X | | | X | | | | | | | | | | | | | | | | | 5 |
| Segurança | | | | | X | | | | X | | X | | | | | X | | X | | | | | | | | 5 |
| Concisão | | X | | | X | X | | | X | | | | | | | | | X | | | | | | | | 5 |
| Detalhamento adequado | | | X | X | X | | | | | | | | | | | X | | | | | | | | | | 4 |
| Unicidade | | X | X | | | | X | | | | | | | | | X | | | | | | | | | | 4 |
| Reputação | | | | | X | | | | X | | | | | | X | | | X | | | | | | | | 4 |
| Objetividade | | X | | | X | | | | X | | | | | | | | | X | | | | | | | | 4 |
| Utilidade | | X | | | X | | | | X | | | | | | | | | | | | | | | | | 3 |
| Privacidade | | | | | | | | X | | | X | | | | | X | | | | | | | | | | 3 |
| Representação apropriada | | X | X | X | | | | | | | | | | | | | | | | | | | | | | 3 |
| Valor agregado | | | | | X | X | | | | | | | | | | | | X | | | | | | | | 3 |
| Validade | | | | X | | | | | | | | | | | | | | | | X | | | | | | 2 |
| Naturalidade | | | X | | | | | | | | | | | | | X | | | | | | | | | | 2 |
| Homogeneidade | | | X | | | | | | | | | | | | | X | | | | | | | | | | 2 |
| Controle de redundância | | | X | | | | | | | | | | | | | X | | | | | | | | | | 2 |
| Consistência estrutural | | | X | | | | | | | | | | | | | X | | | | | | | | | | 2 |
| Portabilidade | | | X | | | | | | | | | | | | | X | | | | | | | | | | 2 |
| Flexibilidade | | | X | | | | | | | | | | | | | X | | | | | | | | | | 2 |
| Tratamento de valores nulos | | | X | | | | | | | | | | | | | X | | | | | | | | | | 2 |
| Armazenamento eficiente | | | X | | | | | | | | | | | | | X | | | | | | | | | | 2 |
| Facilidade de obtenção | | | X | | | | | | | | | | | | | X | | | | | | | | | | 2 |
| Robustez | | | X | | | | | | | | | | | | | X | | | | | | | | | | 2 |
| Ausência de ambigüidades | | | | X | X | | | | | | | | | | | | | | | | | | | | | 2 |
| Registro de apelidos | | | | | | | | | X | | | | | | | | | | | | | | | | | 1 |
| Controle de versões | | | | | | | | | X | | | | | | | | | | | | | | | | | 1 |
| Origem identificada | | | | | | | | | X | | | | | | | | | | | | | | | | | 1 |
| Concordância sobre o uso | | | | | | | | | | | | | | | | X | | | | | | | | | | 1 |
| Integridade | | | | | | | | | | | | | | | | | | | | X | | | | | | 1 |
| Simplicidade | | | | | | | | | | | | | | | | X | | | | | | | | | | 1 |
| Acompanhamento | | | | | | | | | | | | | | | | X | | | | | | | | | | 1 |
| Ubiquidade | | | | | | | | | | | | | | | | X | | | | | | | | | | 1 |
| Custo adequado | | | | | | | | | | | | | | | | X | | | | | | | | | | 1 |

Tabela 2 – Quadro de Referência das propostas de critérios para a qualidade de dados

Legenda:

- 01 - Zmud, 1978
- 02 - Delone e McLean, 1992
- 03 - Redman, 1992
- 04 - Goodhue, 1995
- 05 - Wand e Wang, 1996
- 06 - Wang.,1996
- 07 - Cykana., 1996
- 08 - Matsu Matsumuda e Shouraboura,1996
- 09 - Jarke e Vassiliou, 1997
- 10 - Meyen e Willshire, 1997
- 11 - Kovak., 1997
- 12 - Mandke e Nayar, 1997
- 13 - Gardyn, 1997
- 14 - Ballou e Pazer, 1998
- 15 - Bertolazzi, Scannapieco, 2001
- 16 - Loshin, 2001
- 17 - Wixon e Watson, 2001
- 18 - Kahn, Strong e Wang, 2002
- 19 - Jennex e Olfman, 2002
- 20 - Eckerson, 2002
- 21 - Aubert, 2003
- 22 - Oslen, 2003
- 23 - Shankaranara Yan, Ziad e Wand, 2003
- 24 - Wixon e Todd, 2005
- 25 - Kim, Kishore e Sanders, 2005

2.2.3 Metodologias relacionadas à qualidade de dados

Encontra-se na literatura muitas pesquisas e diferentes enfoques de metodologias relacionadas à qualidade de dados. Estas metodologias diferem na forma como

consideram as seguintes perspectivas: *dimensões e métricas* escolhidas para apoiar os níveis de avaliação da qualidade dos dados; *estratégias e técnicas* adotadas para avaliação e melhoria dos níveis de qualidade dos dados; *fases e passos* que compõem a metodologia; *tipos de custos* associados à qualidade dados, ou seja, custo do processo influenciado por dados errados e custos indiretos gerados pela ausência ou a perda de uma oportunidade em função da má qualidade dos dados; *tipos de dados* considerados; *tipos de sistemas de informação* que utilizam, manipulam, modificam o dado que estão sendo considerados na metodologia; *organizações* envolvidas em processos que criam ou atualizam dados considerados considerando suas estruturas e normas; *processos* que criam ou atualizam dados com o propósito de produzir serviços requeridos pelos usuários considerados; e *serviços* que são produzidos pelos processos considerados.

A tabela 3 apresenta as principais metodologias existentes e em seguida uma breve descrição sobre cada uma dessas metodologias.

| Abrev. | Nome por extenso da Metodologia | Referência |
|-----------|---|---------------------------|
| GQM | Goal Question Metric | Basili, 1994 |
| DWQ | The Data Warehouse Quality Methodology | Jesusfeld et.al. 1998 |
| TDQM | Total Data Quality Management | Wang 1998 |
| TIQM | Total Information Quality Management | English 1999 |
| AIMQ | A methodology for Information Quality assessment | Lee et.al. 2002 |
| DQA | Data Quality Assessment | Pipino et.al. 2002 |
| IQM | Information Quality Measurement | Eppler e Münzenmaier 2002 |
| ISTAT | ISTAT Methodology | Falorsi et.al. 2003 |
| AMEQ | Activity-based Measuring and Evaluating of product information Quality (AMEQ) methodology | Sue e Jin 2004 |
| COLDQ | Cost-effect of Low Data Quality | Loshin 2004 |
| DaQuinCIS | Methodology for the Quatily Assessment of Financial Data | De Amicis e Batini 2004 |
| CIHI | Canadian Institute for Health Information Methodology | Long e Seko 2005 |
| CDQ | Comprehensive methodology for Data Quality Management | Batini e Scannapieco 2006 |

Tabela 3 - Metodologias para qualidade de dados

O GQM (*Goal Question Metric*) baseia-se na suposição de que para se medir de maneira eficaz, alguns objetivos devem ser estabelecidos para que estes sirvam de rota para o estabelecimento de questões que irão orientar a definição de métricas num

contexto particular (BASILI et.al. 1994). No GQM, os objetivos não devem ser avaliados diretamente, mas através de questionamentos que precisam ser respondidos durante a avaliação da qualidade.

A metodologia do TDQM (*Total Data Quality Management*) (WANG, 1998) é resultado de pesquisa acadêmica, porém foi muito utilizada como suporte para iniciativas de reengenharia de dados nas organizações. O objetivo fundamental do TDQM é estender para qualidade de dados os princípios do TQM (*Total Quality Management*) (OAKLAND, 1989).

A metodologia do DWQ (*Data Warehouse Quality*) (JEUSFELD et.al. 1998) estuda as relações entre objetos de qualidade e opções de modelagem no Data Warehousing. Esta metodologia considera conceitos subjetivos de qualidade e provê uma classificação de metas de qualidade conforme definição do grupo de *stakeholders* que procuram essas metas. Eles consideram a diversidade de metas de qualidade existentes e utilizam metadados para defini-las.

A metodologia TIQM (*Total Information Quality Management*) (ENGLISH, 1999) foi concebida para apoiar projetos de *Data Warehouse*. Esta metodologia assume a consolidação das fontes de dados operacionais em uma base única, integrada, utilizada em todos os tipos de agregações efetuadas para a construção do *Data Warehouse*. Essa consolidação elimina erros e fontes de dados heterogêneas. O TIQM foca nas atividades de administração que são responsáveis pela integração de fontes de dados operacionais, elucidando a estratégia que deve ser utilizada pela organização para uma escolha técnica efetiva. Esta metodologia provê uma classificação detalhada de custo-benefício.

A metodologia AIMQ (*A methodology for Information Quality assessment*) (LEE et.al.2002) é a única metodologia em qualidade da informação focada na performance que é um objetivo e uma técnica independente para avaliação da qualidade. Sua fundamentação encontra-se em uma tabela 2X2 (Tabela 4) chamada modelo PSP/IQ classificando dimensões de qualidade conforme sua importância para as perspectivas dos usuários e gestores.

| | | |
|-------------------|---------------------------------|--|
| | Conformidade com especificações | Encontra ou transcende as expectativas do consumidor |
| Qualidade Produto | Informação são | Informação útil |
| Qualidade Serviço | Informação segura | Informação utilizável |

Tabela 4 - PSP/IQ Model (LEE et.al.2002)

O modelo PSP/IQ é a entrada para a metodologia AIMQ cujas fases são resumidas na Figura 1. Nesta metodologia não são providas diretrizes, técnicas, e ferramentas para implementar essas atividades.

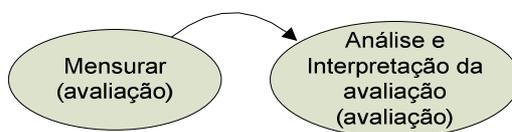


Figura 1: Fases do AIMQ (LEE et.al.2002)

A metodologia CIHI (*Canadian Institute for Health Information*) implementa um método para avaliar e melhorar a qualidade dos dados no Instituto canadense de informações de saúde (LONG e SEKO, 2005). A principal preocupação foi com o tamanho das bases de dados e sua heterogeneidade. A metodologia suporta a seleção de um subconjunto de dados para tratar as fases de avaliação de qualidade e trabalha com uma vasta quantidade de critérios de qualidade para atender a essa heterogeneidade.

A metodologia DQA (*Data Quality Assessment*) (PIPINO et.al. 2002) foi criada com o intuito de prover orientação na definição de métricas de qualidade de dados. Na literatura, as métricas são definidas conforme são necessárias para resolver problemas específicos e por isso são dependentes dos cenários considerados. Essa metodologia é apontada como a primeira iniciativa de identificar princípios métricos de qualidade comuns e generalizados.

O objetivo fundamental da metodologia IQM (*Information Quality Measurement*) (EPPLER e MÜNZENMAIER, 2002) é prover um framework de qualidade de informação apropriado para dados na Web. Utilizando critérios de qualidade, esta metodologia auxilia a seleção e personalização de ferramentas para apoiar o trabalho dos Webmasters na criação, administração e manutenção de sítios Web.

A metodologia ISTAT (*Italian National Bureau of Census*) (ISTAT 2004; FALORSI et.al. 2003) foi criada pela agência nacional italiana de censo para coletar e manter uma elevada qualidade de dados estatísticos referentes a cidadãos e negócios italianos. Seu principal objetivo é definir uma forma de garantir a qualidade de dados integrados originários de diversas fontes locais da administração pública, dividida geograficamente em três níveis: central, regional e periferia. Essa metodologia foca nos tipos de dados mais comuns que são manipulados nos três níveis e que são conhecidos como dados privados; também foca em normas formais desde que estejam priorizando a

regulamentação de atividades de administração de dados de forma que a integração possa satisfazer os requisitos de qualidade.

O principal objetivo da metodologia AMEQ (*Activity-based Measuring and Evaluating of Product information Quality*) (SU e JIN, 2004) é prover uma base rigorosa para a avaliação da qualidade do Produto-Informação (PIQ) e mitigação de riscos alinhados às metas organizacionais. Essa metodologia é específica para avaliar qualidade de dados em empresas de manufatura onde o produto-informação representa o principal componente das bases de dados operacionais. Nas empresas de manufatura, a associação entre produto-informação e processos de produção está diretamente e relativamente padronizada por toda a empresa. Esta apresenta diretrizes para modelagem de informação e processos de produção relacionados.

O objetivo fundamental da metodologia COLDQ (*Cost-Effect of Low Data Quality*) (LOSHIN, 2004) é prover um painel (*scorecard*) para qualidade dos dados apoiado na avaliação do custo-efeito da baixa qualidade dos dados. Também provê um detalhamento da classificação dos custos e benefícios: benefícios diretos são obtidos pela adoção de técnicas de melhoria que evitam custos por se evitar má qualidade dos dados. Sua meta é obter uma avaliação quantitativa da extensão sofrida pelos processos de negócio que são diretamente afetados pela má informação.

A metodologia DaQuinCIS (*Methodology for the Quatily Assessment of Financial Data*) (SCANNAPIECO et.al. 2004) direciona o assunto qualidade de dados para sistemas de informação cooperativos. A origem da cooperação tem dois princípios fundamentais no que se refere à qualidade de dados: primeiro que qualidade de dados é um predicado para a confiança inter-organizacional. Segundo que a má qualidade de dados pode impedir a cooperação e isto poderá gerar diversas consequências. Para tratar do primeiro assunto, a metodologia introduz o conceito de certificação de qualidade de dados, que associa dados com as métricas correspondentes que são trocadas entre organizações junto com os dados. O segundo assunto é tratado pela provisão de um mecanismo de seleção de dados baseado em critérios de qualidade. Esse mecanismo identifica a melhor qualidade dos dados nas bases entre diferentes organizações que utilizam essa cooperação. Desta forma a cooperação é a mola-mestra para melhorar a qualidade.

A metodologia QAFD (*Quality Assessment of Financial Data*) (DE AMICIS e BATINI, 2004) foi concebida para definir padrões de medida de qualidade para dados operacionais financeiros e minimizar os custos com ferramentas de medição de

qualidade. Essa metodologia combina objetivos quantitativos e avaliação subjetiva qualitativa para identificar assuntos de qualidade e selecionar as ações apropriadas para a melhoria da qualidade. Indicadores de dependência de contexto, regras de qualidade de dados, métricas e estratégias para avaliação quantitativa e qualitativa são definidas. De todas as metodologias apresentadas, esta é a única direcionada para dados financeiros.

A metodologia CDQ (*Comprehensive methodology for Data Quality Management*) (BATINI e SCANNEPIECO, 2006; BATINI et.al. 2008) é concebida para ser ao mesmo tempo completa, flexível e simples de aplicação. Perfeição é alcançada considerando técnicas e ferramentas existentes, integrando-as em um framework que pode trabalhar em ambos os contextos: intra-organizacional e inter-organizacional e pode ser aplicada a todos os tipos de dados, estruturados, semi-estruturados e não-estruturados. A metodologia é flexível uma vez que suporta o usuário na seleção de técnicas mais satisfatórias e ferramentas dentro de cada fase e contexto.

CDQ é simples uma vez que é organizado em fases e cada fase é caracterizada por uma meta específica e um conjunto de técnicas para aplicar. Esta metodologia inova no fato de apoiar a seleção do melhor processo de avaliação de qualidade que maximize os benefícios dentro dos limites financeiros definidos. CDQ enfatiza a fase de elucidação das exigências iniciais. De fato, as outras metodologias implicitamente assumem que um conhecimento contextual foi previamente anexado e modelado. O foco é em como alcançar qualidade de dados total sem prover indicações de como utilizar o conhecimento prévio. A meta do CDQ é obter uma avaliação quantitativa da extensão na qual os processos de negócio são afetados pela má qualidade da informação.

Para que seja possível então, um maior êxito nos projetos de BI e obtenção de máxima qualidade de dados na organização é necessário a adoção de um processo de Governança de Dados. Com isso na próxima seção descreve-se uma breve conceituação e as três propostas de Governança de Dados existentes, oriundas de fornecedores de solução de TI, sendo duas proposições feitas por fornecedores de hardware e software.

2.3 Governança de Dados

Governança, de forma genérica, refere-se à forma como a organização age para garantir que as estratégias sejam implementadas, monitoradas e alcançadas (RAU, 2004). Mais

especificamente, Governança de Dados corporativos está relacionada à forma de tratar o dado como um recurso da organização. Isso envolve processos, políticas, padronização, organização e tecnologias necessárias para manipular e assegurar a disponibilidade, acessibilidade, qualidade, consistência, auditabilidade e segurança dos dados na empresa ou instituição.

Um *framework* de Governança de Dados pode ajudar às organizações a manipularem os dados de forma mais eficiente. Esse *framework* deve prover definições consistentes, estabelecer uma administração de dados na organização e mensurar e rastrear a qualidade dos dados transacionais e analíticos utilizados ao longo da organização. O custo da informação pode diminuir com a implantação efetiva de um *framework* de Governança de Dados. Bases dados duplicadas espalhadas pela organização podem ser eliminadas e o custo com processos de limpeza dados podem ser reduzidos com a existência de fontes de dados com mais qualidade. Uma efetiva Governança de Dados auxilia a melhorar os esforços de controle e conformidade aos padrões da empresa (*compliance*).

Fatores críticos de sucesso para Governança de Dados (FISHER, 2009):

1. Definição clara e sucinta do proprietário: se a ninguém pertence o esforço de Governança de Dados, este tenderá a ser negligenciado e a sofrer com a ausência de propósitos claros. O assunto Governança de Dados tem que ser tratado como um assunto organizacional. Se somente a TI carrega essa bandeira, a área de negócio pode perceber que as iniciativas de governança não atendem a suas necessidades ou não requer nenhum trabalho ou esforço por parte deles. Um grupo ou comitê de Governança de Dados deve ser organizado. Pode ser composto de vários representantes de áreas funcionais da organização. Este comitê deve definir as diretrizes e políticas de tratamento dos dados, procedimentos e padrões para a empresa como um todo. Deve ser um tratamento uniforme e válido para toda a organização e atualizado mediante consenso quando necessário.
2. Reconhecimento do valor agregado: é difícil quantificar o valor do dado em termos monetários, mas é fato que o conjunto de dados é um dos recursos mais importante de qualquer negócio. Sem padrões de dados e qualidade na informação de suporte à decisão, os negócios não funcionam bem, podendo implicar em prejuízo para a organização.

3. Políticas de dados e procedimentos: para as empresas serem mais eficientes, políticas de dados e procedimentos devem ser adotados por todas as unidades de negócio e aplicáveis ao negócio como um todo. É essencial ter um modelo de negócio comum e é fundamental que a organização inteira possa confiar e seguir este modelo. Senão, o caos dos dados reinará e problemas de qualidade de dados podem crescer e se tornar intratáveis.

A seguir é apresentado um resumo das propostas de Governança de Dados encontradas na literatura: a primeira da Oracle, a segunda da IBM e por último a do Data Governance Institute (DGI).

2.3.1 Programa de Governança de Dados proposto pela Oracle

A Oracle defende que o sucesso de um programa de Governança de Dados se apóia na confiança em vários requisitos, mesmo que a iniciativa de governança se propague pela empresa toda ou por uma divisão ou área de negócio. A Oracle define seis partes para o seu programa de Governança de Dados (Figura 2): *Estratégia, Organização, Políticas e Processos, Medidas e Monitoramento, Tecnologia e Comunicação*. (HSU, V. et.al., 2009)



Figura 2 - Governança Dados da Oracle (HSU, V. et.al., 2009)

Na *Estratégia* se articulam a visão, missão, objetivos e princípios norteadores de todo o programa. A missão deve comunicar o propósito básico do programa de governança e atua como um referencial para todas as decisões futuras. Serve para alinhar os objetivos do programa à estratégia corporativa e alinhar pessoas, processos e tecnologia aos objetivos comuns. Na *Estratégia* também se define o escopo para os

padrões de Governança de Dados, papéis e responsabilidades. A *Estratégia* inclui um modelo de operação, constituído pela estrutura organizacional de Governança de Dados; o modelo deve garantir a visão plena e contemplar as ações para resolver os desafios relacionados a dados conforme surgirem. Deve contemplar definições tais como: Quem é o dono e o responsável, direito de decisão e controles, ou seja, quem tem autoridade para adicionar novos registros; que informações são necessárias de serem verificadas e por quais processos; por quanto tempo o dado tem que ser arquivado e retido; e qual deve ser o processo de auditoria. Também devem ser definidas as métricas que a organização deseja medir, durante os processos de mensuração e monitoramento.

Quanto à *Organização* de Governança de Dados, a Oracle defende a formação de dois grupos: um comitê líder e um comitê operacional de Governança de Dados. O comitê líder é formado por gestores seniores a nível da presidência e vice-presidência das diversas áreas da empresa. Este grupo é um facilitador, de forma que as ações de Governança de Dados se propaguem pelas áreas de negócio e por toda empresa. O comitê operacional é o que está envolvido com as análises e o trabalho propriamente dito. Os membros do programa são formados a partir deste comitê, com o apoio de outros profissionais quando necessário, e que tenham perfil profissional para desempenhar funções como estrategistas de dados, guardiães de dados, e orientadores para modelagem de dados e para a Governança de Dados. Esses papéis e responsabilidades devem ser devidamente documentados e comunicados em toda a empresa.

As *Políticas e Processos* de Governança de Dados são diretrizes documentadas, processos e tarefas que provêm a direção para o administrador de dados e outros administradores. São as ferramentas que apóiam o trabalho de qualidade de dados, a definição de padrões da empresa e regras confiáveis. As políticas devem se referir a:

1. Administração do ativo-dado;
2. Padronização de nomes;
3. Acesso, segurança e proteção dos dados;
4. Regras de equiparação e fusão dos dados;
5. Resolução de conflitos e princípios de autorização.

Quanto à *Mensuração e Monitoramento*, é a parte que visa mostrar a aderência das políticas e processos e encorajar ao contínuo melhoramento da administração dos recursos, mostrando continuamente quão importante e valioso é o programa para a

empresa. Ferramentas de *data profiling* e qualidade de dados provêm indicadores de performance que os gestores da empresa julgam importante.

A proposta da Oracle é baseada na tecnologia MDM (*Master Data Management*) que provê as fundações tecnológicas para a Governança de Dados. Segundo a Oracle, MDM engloba qualidade de dados, estrutura de dados, arquitetura de dados, administração de dados corporativos e segurança de dados. Na parte de *Tecnologia e Comunicação*, uma estratégia de comunicação deve ser bem delineada, definindo-se quem deve ser comunicado, que informações são importantes e que frequência deve ser feita a atualização da informação, e qual o método mais eficiente de disseminação da informação na empresa. A comunicação do andamento do programa de Governança de Dados deve ter a definição de um processo de distribuição formal e ser disponibilizado em um portal da empresa ou na intranet ou *newsletter* de forma que todos tenham acesso às atualizações, métricas e realizações.

A Oracle defende que Governança de Dados é a definição das seis partes descritas anteriormente, e propõe uma solução chamada Oracle Data Governance Manager como solução tecnológica para apoiar a TI na administração dos dados e qualidade dos dados.

2.3.2 Programa de Governança de Dados proposta pela IBM

A proposta de Governança de Dados da IBM (IBM DATA GOVERNANCE COUNCIL MATURITY MODEL, 2007) é o resultado de diversas discussões e benchmarks entre membros da IBM, parceiros de negócio, universidades e 55 companhias líderes de mercado. Destas discussões surgem os elementos de Governança de Dados (Figura 3), composta de onze disciplinas essenciais agrupadas em quatro grupos: *Efeito/Resultado*, *Autorização*, *Disciplinas essenciais* e *Disciplinas de apoio*.

Do grupo *Efeito/Resultado*, fazem parte as disciplinas:

- *Administração de riscos de dados e Conformidade*, que define a metodologia pela qual os riscos são identificados, qualificados, quantificados, evitados, aceitos ou mitigados.
- *Valor criado (ou agregado)*, processo pelo qual os ativos relativos a dados são qualificados e quantificados para permitir ao negócio maximizar o valor criado.

Do grupo *Autorização*, fazem parte as disciplinas:

- *Estrutura e consciência Organizacional*, que descreve o nível de responsabilidade entre negócio e TI e reconhecimento da responsabilidade fiduciária da Governança de Dados nos diferentes níveis da administração.
- *Políticas*, que é a articulação por escrito do comportamento organizacional desejado em relação aos dados obtidos e produzidos.
- *Administração*, que é a disciplina de controle de qualidade definida para assegurar o cuidado do dado custodiante de forma a aumentar o valor do dado como ativo, o controle organizacional e mitigar risco.

Do grupo *Disciplinas essenciais*, fazem parte as disciplinas:

- *Qualidade de Dados*, que define métodos para mensurar, aprimorar e certificar a qualidade e integridade da produção, teste e arquivamento de dados.
- *Administração do ciclo de vida da informação*, política sistemática com enfoque na coleção, utilização, retenção, e deleção da informação.
- *Segurança e Privacidade*, onde se descreve políticas, práticas e controles utilizados pela organização para mitigar risco e proteger o ativo-dado.

Do grupo *Disciplinas de apoio*, fazem parte as disciplinas:

- *Arquitetura de Dados*, onde se trata o desenho arquitetural de dados estruturados e não-estruturados utilizados nos sistemas e aplicações, possibilitando a disponibilização dos dados e distribuição para os usuários apropriados.
- *Classificação / Metadados*, que trata dos métodos e ferramentas utilizadas para criação de definições semânticas comuns para termos de negócio e TI, modelos de dados e repositórios. Metadados que constrói entendimento humano e computacional.
- *Auditoria e Relatório*, onde se definem os processos organizacionais para monitoramento e mensuração do valor dos dados, riscos e eficácia da governança.



Figura 3 - Elementos de uma efetiva Governança Dados segundo a IBM (IBM DATA GOVERNANCE COUNCIL MATURITY MODEL, 2007)

2.3.3 Programa de Governança de Dados segundo o Data Governance Institute

A proposta do DGI – Data Governance Institute (GWEN T., 2009), é definir um framework prático e ágil que auxilie os usuários-chave (*stakeholders*) das organizações a entender de forma clara e objetiva o propósito de um programa de governança e administração de dados. O framework é formado por 10 componentes, reunidos em 3 grupos:

O primeiro grupo intitulado: *Regras Gerais e de Entendimento, cujos componentes são:*

1. *Missão e Visão da Governança de Dados;*
2. *Metas, Métricas de governança e Mensuração de êxito e fundamentação estratégica;*
3. *Regras e definições de dados;*
4. *Direitos de decisão;*
5. *Responsabilidades;*
6. *Controles.*

O segundo grupo chama-se: *Pessoas e Corpo Organizacional*, cujos componentes são:

1. *Stakeholders de dados*;
2. *Escritório de Governança de Dados (DGO)*;
3. *Administrador de Dados*.

E o terceiro grupo chamado Processos, com um único componente: *Processos de Governança de Dados contínuo, pró-ativo e reativo*.

O DGI ainda define um ciclo de execução do programa e Governança de Dados, composto dos seguintes passos:

1. Desenvolvimento valores de negócio
2. Preparar um roteiro
3. Planejar e fundamentar
4. Desenhar o programa
5. Divulgar o programa
6. Governar os dados
7. Monitorar, medir e relatar

Segundo o DGI os gestores das organizações somente percebem a importância de um programa de Governança de Dados se esses esforços auxiliarem a organização a lucrar mais, alcançar sua missão ou aumentar o valor dos seus recursos; gerenciar melhor os custos e complexidade e apoiar outras iniciativas, tais como, segurança, ou conformidade (“*compliance*”) e privacidade.

O framework está organizado utilizando o padrão:

QUEM-O QUE-QUANDO-ONDE-PORQUE.

Com isso o framework define:

- Porque seu programa deve existir.
- O que estará realizando.
- Quem estará envolvido nas tarefas, junto com suas responsabilidades específicas.
- Como eles estarão trabalhando juntos para prover valor para a organização.
- Quando eles irão executar processos específicos.

Outra forma de ver o programa é considerar:

- As regras que o programa estará criando, coletando, alinhando e formalizando (políticas, requisitos, padrões, responsabilidades, controles, definições de dados, etc.) e as regras de compromisso que descrevem como diferentes grupos trabalham juntos na confecção dessas regras e implementação.
- As pessoas e estrutura organizacional envolvidas na confecção e implementação dessas regras e
- Os processos que essas pessoas utilizam para governar dados, enquanto criam valor, administram custo e complexidade e garantem conformidade.

Na próxima seção é explorado os métodos de desenvolvimento de arquitetura empresarial para melhor entendimento e contextualização da proposta apresentada neste trabalho.

2.4 Arquitetura Empresarial

Em uma organização, os componentes do negócio precisam funcionar de forma integrada e sincronizada. Eles interagem entre si permanentemente e o mau funcionamento de qualquer componente, afeta a todos os outros componentes da estrutura produtiva.

Arquitetura Empresarial é um plano que "age como uma força de colaboração" entre os aspectos de planejamento de negócios (metas, visões, estratégias e princípios de governança), aspectos das operações de negócio (termos de negócios, estruturas organizacionais, processos e dados), aspectos da automação (sistemas de informação e bases de dados) e infra-estruturas tecnológicas (computadores, sistemas operacionais e redes) que viabilizam o negócio (SCHEKKERMAN, 2008). Ou seja, a Arquitetura Empresarial é um grande modelo que documenta e representa todos os comportamentos que fazem uma organização funcionar. Deve contemplar: os dados que são processados, quem faz isso, onde as coisas se encontram e porque tudo isso é feito. A crescente preocupação com os riscos operacionais e a legislação vigente tem forçado as empresas a implementá-la. Existem três razões básicas para que a Arquitetura Empresarial seja considerada importante e faça diferença no contexto corporativo (BPM-ADVISOR, 2010):

1. Reduz as incertezas;

2. influencia o comportamento corporativo relacionado aos processos de negócio;
3. utiliza tecnologia da informação, trazendo grande valor agregado ao negócio.

Pode-se destacar cinco benefícios da arquitetura empresarial (BPM-ADVISOR, 2010):

- Garante que os empreendimentos realizados estejam alinhados com os objetivos gerenciais.
- Garante que as regras de negócios utilizadas sejam consistentes, que os dados e seus usos sejam imutáveis, que as interfaces e os fluxos de informação sejam padronizados, e que a conectividade e a operabilidade da organização estejam gerenciados.
- Facilita e gerencia as mudanças para qualquer aspecto da organização.
- Reduz a quantidade de sistemas desenvolvidos, a geração de aplicações, as janelas de tempo da modernização e a necessidade de recursos.
- Reduz o esforço para a manutenção dos produtos do portfólio de TI, mantendo conformidade com os modelos técnicos de referência especificados.

Pode-se encontrar na literatura várias propostas tanto de frameworks de arquitetura empresarial proprietários quanto públicos. Estes frameworks encontram-se detalhados no ANEXO IV deste trabalho. Aqui são destacados os Frameworks: Zachman (ZACHMAN, 2009) e TOGAF (THE OPEN GROUP, 2009), muito utilizados em projetos de arquitetura empresarial. Eles são descritos de forma resumida, o que possibilitará uma melhor compreensão da proposta aqui apresentada.

2.4.1 Framework Zachman

O “Framework” Zachman (ZACHMAN, 2009) é na verdade uma taxonomia para a organização de artefatos arquiteturais, ou seja, documentos de projeto, especificações e modelos, que considera a quem se destina o artefato (por exemplo, proprietário ou construtor do negócio) e qual problema específico (por exemplo, dados e funcionalidade) está sendo abordado. Como o próprio ZACHMAN descreveu em seu trabalho: “O framework de arquitetura corporativa, conforme aplicado às empresas é simplesmente uma estrutura lógica para classificação e organização das representações descritivas de uma empresa, significativa à administração da empresa, assim como ao desenvolvimento dos sistemas corporativos (ZACHMAN, 2009).

Zachman explicou sua taxonomia utilizando como analogia, o setor de construção civil. Nesse setor, os artefatos arquiteturais estão implicitamente sistematizados por meio de uma organização bidimensional. Uma dimensão representa os vários “participantes do jogo”. Em um edifício físico, alguns desses participantes são o proprietário (o que paga o projeto), o construtor (coordena toda a construção) e um conselho de zoneamento (que assegura o cumprimento pelo construtor das normas locais de edificações). Um arquiteto de edificações elabora vários artefatos para cada um desses participantes. Cada participante exige informações completas, mas o que constitui a completude varia de acordo com o participante. O proprietário está interessado na descrição completa da funcionalidade e da estética do edifício. O construtor está interessado em uma descrição completa dos materiais e do processo de construção. O proprietário não se importa com o posicionamento das estacas das paredes. O construtor não se importa com o alinhamento das janelas do dormitório em relação ao sol da manhã.

O Framework Zachman não é uma metodologia para criar uma aplicação (uma instância) do objeto. O Framework Zachman é a ontologia para descrever a organização. Um framework é uma estrutura, enquanto que esta metodologia é um processo. O Framework Zachman é uma ontologia - uma teoria da existência de um conjunto estruturado de componentes essenciais de um objeto para o qual expressões explícitas são necessárias e talvez até mesmo obrigatórias para criar, operar e manter o objeto (o objeto pode ser uma empresa, um departamento, uma cadeia de valor, uma solução, um projeto, um avião, uma construção, um produto, uma profissão, etc).

Atualmente o Framework Zachman é retratado como uma matriz 6 x 6 com os questionamentos de comunicação representados nas colunas e nas linhas constam os diferentes pontos de vista e níveis de detalhe relativos à informação. As classificações do framework são representadas pelas células, ou seja, a interseção entre as linhas e colunas, representando o aspecto e ponto de vista correspondente. Esta matriz necessariamente constitui o conjunto total de representações descritivas que são relevantes para a descrição de uma empresa em particular. Cada perspectiva traduz uma visão da organização adequada a um determinado fim ou atividade (ZACHMAN, 2009).

John Zachman (2009) considera 6 perspectivas no seu framework (Figura 4):

1. Escopo e o contexto, representando o ponto de vista do estrategista como teorizador sobre a organização. Normalmente, contém informação relevante para

o planejamento estratégico de alto nível e, é claro, o próprio conteúdo da Estratégia da organização.

2. Conceitos de negócio, representando a visão da liderança executiva (vistos como proprietários dos processos de negócio e informações relacionadas). Contém tipicamente descrição detalhada da organização no nível de processos de negócio.
3. Sistemas de informação (nível lógico), com a visão dos arquitetos de sistemas (designers).
4. Informações sobre a infraestrutura tecnológica (nível físico) da organização, sendo o ponto de vista dos engenheiros enquanto construtores.
5. Refere-se à descrição dos componentes que a organização utiliza para operar, sendo a visão dos técnicos implementadores.
6. Representa as operações propriamente ditas da organização, instanciadas pelos seus colaboradores participantes.

As colunas correspondem às clássicas perguntas 5W1H (What / Who / Where / When / Why / How) aplicadas à organização (ZACHMAN, 2009). As colunas, portanto, referem-se aos diferentes aspectos da organização que precisam ser conhecidos. A dimensão “o quê” refere-se à informação que a organização precisa de ter e gerir, o “como” refere-se aos processos e atividades da organização, o “quem” refere-se aos departamentos, pessoas, perfis e competências da organização, o “quando” refere-se ao acontecimentos a que a organização deve reagir, o “porquê” justifica a razão de ser de cada elemento que constitui a organização e finalmente o “onde” traduz os locais onde esses elementos residem. Por exemplo, ao nível estratégico, o “onde” indica a cidade ou país em que a organização deve estar representada. Ao nível do negócio, o “onde” indica onde é que a informação reside, os processos de negócio são concretizados, onde é que as pessoas trabalham, etc.

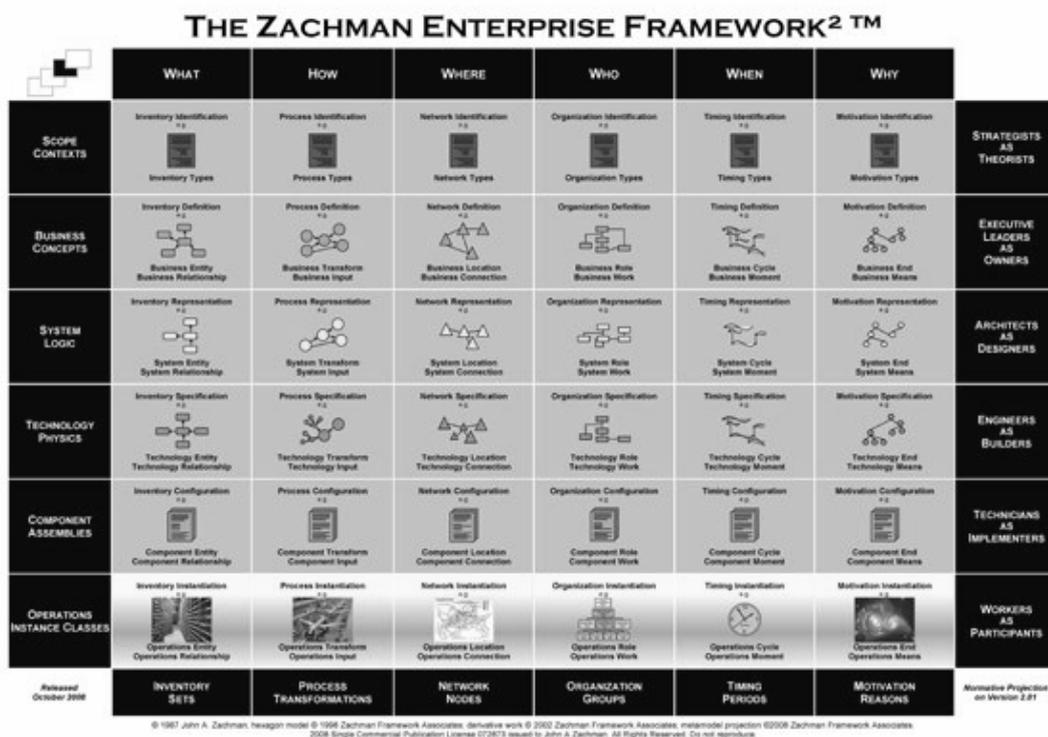


Figura 4 - Zachman Framework (THE OPEN GROUP, 2008)

A matriz resultante possui 36 células: cada célula tem um propósito bem definido, e facilmente compreensível pelas várias áreas da organização. O resultado dessa combinação é uma matriz suficientemente genérica para poder ser usada como repositório das várias descrições da organização, bem como evidenciar os seus desalinhamentos.

2.4.2 Framework TOGAF

The Open Group Architecture Framework (THE OPEN GROUP, 2009) é um método detalhado e um conjunto de ferramentas de suporte para o desenvolvimento de uma arquitetura empresarial. Ele pode ser utilizado gratuitamente para uso interno, por qualquer organização. O TOGAF foi desenvolvido por membros do *The Open Group*, trabalhando em parceria com o *Architecture Fórum* (THE OPEN GROUP, 2009). O desenvolvimento da 1ª. Versão foi em 1995 baseado no “*Technical Architecture Framework for Information Management* (TAFIM)”, desenvolvido pelo *Department of Defense* (DoD) dos Estados Unidos. O DoD deu ao *The Open Group* permissão e encorajamento para criar o TOGAF que foi o resultado de muitos anos de esforço de desenvolvimento e muito investimento por parte do Governo dos Estados Unidos. O

TOGAF reúne as boas práticas de modelagem organizacional sendo composto de um método detalhado e um conjunto de ferramentas de suporte para o desenvolvimento de arquiteturas organizacionais.



Figura 5 - Arquitetura Corporativa do TOGAF (THE OPEN GROUP, 2009)

Conforme apresentado na Figura 4, o TOGAF divide uma arquitetura corporativa em quatro categorias, como segue:

1. Arquitetura de negócio, que descreve os processos que o negócio usa para cumprir suas metas;
2. Arquitetura de aplicativos, que descreve como aplicativos específicos são programados e como interagem;
3. Arquitetura de dados, que descreve como os armazenamentos de dados são organizados e acessados;
4. Arquitetura técnica, onde as infra-estruturas de hardware e software que suportam os aplicativos e suas interações são definidas e documentadas.

O TOGAF descreve a si próprio como um *framework* corporativo, mas a parte mais importante do TOGAF é o método de desenvolvimento da arquitetura, mais conhecido como ADM. O ADM é uma metodologia para a criação da arquitetura. É uma sequência iterativa de passos para desenvolver uma arquitetura empresarial completa. O TOGAF observa o mundo da arquitetura corporativa como um conjunto de arquiteturas, variando de extremamente genérico para altamente específico. O ADM do TOGAF oferece um processo para encaminhar as ações do genérico para o específico, como se pode observar na Figura 6:

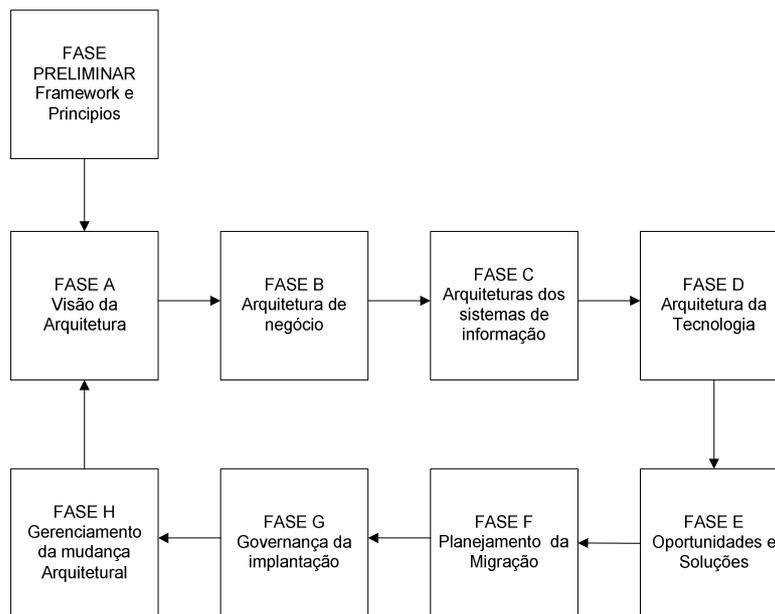


Figura 6 - Método de desenvolvimento da arquitetura (ADM) do TOGAF (THE OPEN GROUP, 2009)

A Fase Preliminar prepara a organização para uma implementação de sucesso. Isto significa estudar e entender o ambiente de negócio, estabelecer compromissos nos níveis gerenciais mais elevados, definir acordos sobre o escopo do trabalho, estabelecer os princípios para o projeto e os resultados, estabelecer a estrutura de governança e definir acordos sobre métodos a serem adotados.

A fase A – Visão da Arquitetura inicia uma interação dos processos da arquitetura e é requerida a cada início de evolução da arquitetura, determinando escopo, restrições, expectativas e validando o contexto dos negócios.

Já a fase B – Arquitetura de Negócio é onde se cria as fundações para um empreendimento empresarial organizado, mapeando: os processos de negócios e pessoas; seus relacionamentos entre cada um e o ambiente; e os princípios que governam seus modelos e sua evolução. Ou seja, nesta fase mostra-se como a organização endereça seus objetivos de negócio.

Quanto à fase C – Arquitetura dos sistemas de informação, é a fase que mostra como a TI endereça os objetivos de negócio da companhia. Criam-se as fundações para uma estrutura de TI organizada, apoiada em definições dos relacionamentos entre cada um e o ambiente e dos princípios que governam seus modelos e evolução.

Na fase D – Arquitetura da tecnologia é onde se cria o fundamental para uma infra-estrutura de TI organizada, ou seja, definem-se as tecnologias de hardware, software e comunicações.

A fase E é utilizada para identificação dos maiores projetos de implementação. Nesta fase é necessário decidir sobre fazer ou comprar ou reutilizar; se será utilizado *outsourcing* e terceirização de serviços ; avaliação dos custos; se vai se utilizar software de código aberto; define-se prioridade dos ativos e identificam-se dependências.

Para os projetos identificados na Fase E, na Fase F – Planejamento da migração, deverá se realizar análise de custos e benefícios e avaliação de riscos, produzindo-se assim um roteiro (*Road map*) da implementação.

Chegando à Fase G – Governança da Implementação, cabe definir as restrições da arquitetura nos projetos de implementação, bem como os contratos internos e níveis de serviço da arquitetura e monitorar o trabalho de implementação para garantir conformidade.

Por último, a Fase H – Gerenciamento de mudança arquitetural garante que as alterações na arquitetura são gerenciadas de uma forma coesa e projetada. Define e apóia a arquitetura empresarial a proporcionar a flexibilidade necessária para evoluir rapidamente, em resposta às mudanças na tecnologia ou no ambiente empresarial.

2.5 Considerações sobre Arquitetura Corporativa

A arquitetura corporativa é um caminho e não terá validade se não trazer valor real ao negócio, tão logo quanto possível. Um dos objetivos mais importantes de qualquer arquitetura corporativa é fazer com que as duas áreas, negócio e tecnológica, juntem-se para que ambas trabalhem de modo eficaz para a realização das mesmas metas. Por este motivo optou-se por utilizar o TOGAF que possui uma metodologia de desenvolvimento de arquitetura empresarial (ADM) que proporciona uma orientação ao usuário através de um processo, um passo a passo, para a criação de uma arquitetura corporativa.

Esta é bastante útil para o entendimento e criação de um modelo de governança efetiva para a arquitetura corporativa. Também ajuda o usuário a assimilar o conceito de arquitetura corporativa na organização e desenvolver uma cultura na qual promova sua valorização e utilização.

Outra vantagem importante de se destacar é o fato de que o TOGAF proporciona um baixo comprometimento com fornecedores, ou seja, a probabilidade do usuário ficar preso a uma organização de consultoria específica pela sua adoção é nula. Também se pode considerar uma vantagem a quantidade razoável e qualidade de informações gratuitas e acessíveis que se pode encontrar sobre essa metodologia.

Apesar do framework Zachman também ser muito utilizado, ele não nos oferece um processo, passo a passo, para criar uma nova arquitetura. Além disso, não ajuda muito a decidir se a arquitetura futura que está sendo criada será a melhor arquitetura possível. Também não oferece uma abordagem para mostrar a necessidade de uma arquitetura futura. Pelas vantagens destacadas anteriormente e consideradas fundamentais para o desenvolvimento de um framework de Governança de Dados, optou-se por utilizar o TOGAF-ADM.

Conforme se pode observar, pelo que foi explanado anteriormente, encontra-se na literatura diversas pesquisas e diferentes enfoques, no que se refere à qualidade de dados. Também é possível identificar variâncias nas propostas de Governança de Dados encontradas até o presente momento. Inicialmente, por acreditar que uma sistemática para aferir e melhorar a qualidade dos dados nas organizações era necessária, apoiando assim as equipes de qualidade de dados, de desenvolvimento e usuários finais, aplicou-se a pesquisa que se encontra descrita no próximo capítulo. A partir das conclusões obtidas da compilação do resultado desta pesquisa, foi que se desenvolveu o framework de Governança de Dados descrito no capítulo 4, utilizando o TOGAF-ADM pelas razões aqui descritas.

3. Pesquisa Exploratória: Necessidade de apoiar equipe de TI na explicitação da qualidade dos dados em DW e/ou DM (Panorama Brasil)

Na fase inicial deste trabalho acreditava-se que após a reunião de diversos conceitos existentes na literatura sobre qualidade de dados, seria interessante a elaboração de um framework conceitual que reunisse essas concepções em um conjunto de classes. Esta sistemática poderia ser utilizada para apoiar as equipes de TI na aferição e melhoria da qualidade dos dados nas organizações, principalmente no que se refere a dados gerenciais e facilitaria o trabalho das equipes de qualidade de dados, equipes de desenvolvimento do DW e usuários finais.

A partir desta perspectiva, surgiu então o interesse em se verificar como os profissionais de TI (mais precisamente os envolvidos em dados para apoio à decisão) estavam lidando com a questão da qualidade dos dados. Para tal, aplicou-se um questionário online com o intuito de avaliar como a qualidade da informação estava sendo tratada nas organizações, e também avaliar a necessidade de apoiar equipes de TI na explicitação de critérios de qualidade nos dados (especificamente em Data Warehouse e/ou Data Marts - DMs).

Com base nas respostas deste questionário, observou-se que mais de 80% desses profissionais, que fazem tratamento e limpeza dos dados no processo de ETL (extração transformação e carga), acreditavam garantir de 80 a 100% da qualidade de dados, porém 60% deles afirmavam que os usuários do DW e/ou DM continuam manifestando necessidade de ter mais informações sobre a qualidade do conteúdo do mesmo. As ações de melhoria na qualidade de dados não são tão facilmente percebidas. Mais de 60%, dos respondentes não conseguiam identificar que pessoas da administração patrocinam projetos de melhoria contínua na qualidade dos dados ou mesmo que estivessem envolvidas em ações de melhoria na qualidade dos dados

Neste capítulo serão discutidos a metodologia, o contexto e a forma de aplicação do questionário desta pesquisa. Também se encontram descritos os resultados obtidos a partir da aplicação deste questionário, bem como o perfil de seus questionados, com o intuito de responder às seguintes questões: como os profissionais de TI estão lidando com a questão da qualidade dos dados, principalmente os que estão envolvidos em projetos de apoio à decisão e se existe a necessidade de apoiar essas equipes em atividades relacionadas à gerência da qualidade e explicitação de critérios de qualidade nos dados.

3.1 Considerações Metodológicas

Após estudo pormenorizado das propostas existentes na literatura sobre qualidade de dados, em particular em DW, esta pesquisa surgiu do interesse em se verificar como os profissionais de Tecnologia da Informação (TI), mais precisamente os envolvidos em projetos de DW e/ou Data Marts (DM), que estão envolvidos em dados de apoio à decisão, lidam com a questão da qualidade dos dados. O intuito desta pesquisa exploratória foi o de avaliar como a qualidade da informação está sendo tratada nas organizações, principalmente no que se refere a informações para apoio à decisão. Também se quis avaliar a necessidade de apoiar as equipes de TI nas atividades relacionadas à gerência da qualidade, incluindo a explicitação de critérios de qualidade nos dados, especificamente em *Data Warehouse* e/ou *Data Marts*.

Como procedimento metodológico adotado foi utilizada a *pesquisa exploratória*, pois esta vai além da descrição das características, analisando e explicando por que ou como os fatos estão acontecendo. As pesquisas exploratórias, segundo Gil (1995), são desenvolvidas com objetivo de proporcionar visão geral, de tipo aproximativo, acerca de determinado fato. VERGARA (2004) afirma que uma investigação exploratória é realizada em área na qual há pouco conhecimento acumulado e sistematizado. ANDRADE (2001) complementa dizendo que esta se configura como a fase preliminar, que busca proporcionar maiores informações sobre o assunto que vai se investigar. Pesquisa exploratória visa proporcionar maior familiaridade com o problema com vistas a torná-lo explícito ou a construir hipóteses. Esta modalidade de pesquisa envolve levantamento bibliográfico, entrevistas com pessoas que tiveram experiências práticas com o problema pesquisado, e análise de exemplos que estimulem a compreensão.

Foi utilizado um questionário online de múltiplas escolhas fechadas com uma série de respostas possíveis (Anexo I), porém permitindo que o respondente pudesse estar complementando a resposta acrescentando algo relevante. Este foi construído em dois blocos temáticos: qualidade de dados e qualidade de dados em DW ou DM. Inicialmente, este questionário ficou disponível por 40 dias, porém este prazo foi estendido por mais 30 dias em função de ter sido disponibilizado em um período de festas Natalinas e férias escolares (muitos profissionais tiram férias em janeiro para viajar com a família).

O contato inicial foi feito via email para 60 profissionais de TI de organizações de diversos ramos de negócio, sediadas no Rio de Janeiro, e que possuem algum relacionamento com o pesquisador (profissionais com quem havia trabalhado anteriormente ou que conheceu em eventos ou treinamentos). Alguns declinaram da participação, porém houve colaboração no sentido de indicarem outros profissionais envolvidos em projetos de apoio a decisão e que poderiam estar participando mais efetivamente da pesquisa. Os principais motivos alegados pelos que declinaram foram: pelo fato das organizações onde se encontram ainda considerarem o tema qualidade de dados como um assunto delicado; pelo fato de não possuir nenhum projeto e nem ferramentas de qualidade de dados; por acreditarem que ferramentas de qualidade de dados ainda são caras e que neste momento a organização se encontra em fase de estudo para construção de DW e/ou DM.

Esta pesquisa exploratória foi respondida por 43 profissionais de TI, porém já não mais restrita a empresas no Rio de Janeiro. Após a compilação das respostas e da análise de alguns comentários efetuados, percebeu-se que o assunto transcende a área de TI, devendo ser tratado por toda organização.

3.2 Contexto da pesquisa

Qualidade de dados é crucial para os trabalhadores do conhecimento e gestores terem confiança nas fontes de dados e no que diz respeito a informações gerenciais torna-se uma necessidade crítica. Mas são poucas as empresas que podem confiar totalmente em seus dados (CAMPOS, 2005). Para analisar como uma organização adota práticas de gestão de dados, há uma série de perguntas a serem elaboradas. Uma delas é se a empresa tem problemas, custos adicionais ou perdas financeiras, ficando exposta a riscos em decorrência da baixa qualidade de seus dados. Outra, por exemplo, é o nível

de dependência que a empresa tem de processos automatizados de tomada de decisão. São também considerados indicativos do comprometimento com gerenciamento de dados a atenção que a alta administração dá ao tratamento de dados e a existência de uma estratégia para o assunto (CAMPOS, 2005).

O questionário online foi disponibilizado a profissionais de TI com vínculo empregatício em organizações sediadas no Brasil e teve por finalidade avaliar como está sendo tratada a questão da qualidade da informação utilizada para apoio à decisão e a necessidade de apoiar equipes de TI nas atividades relacionadas à gerência da qualidade, incluindo a explicitação de critérios de qualidade nos dados, especificamente em *Data Warehouse* e/ou *Data Marts*.

O período de coleta de dados ocorreu de 1º de dezembro de 2009 a 12 de fevereiro de 2010, sendo enviado um *e-mail* ao potencial respondente com *link* desenvolvido utilizando-se o aplicativo existente no site *Survey* (<http://www.keysurvey.com/survey/262501/1a31>). Tal e-mail foi encaminhado após um primeiro contato, também por email, explicando o objetivo da pesquisa e solicitando a autorização e colaboração do mesmo.

Apesar de a premissa estar na possibilidade de uma sistemática para implantação de mecanismos de quantificação, qualificação da qualidade de dados, bem como sua avaliação e auditoria, e apoio às equipes de TI, o resultado desta pesquisa exploratória mostrou que outros problemas influenciam na implantação de projetos de qualidade de dados na empresa, incluindo desconhecimento da complexidade do assunto e problemas culturais e administrativos.

3.3 Questionário e forma de aplicação

A coleta de dados foi feita por meio de um questionário online (Anexo I) disponibilizado na Web. A escolha desse instrumento de pesquisa justifica-se pela conveniência para o entrevistado em responder no momento e local que julgar mais apropriado, não tendo influência de opiniões e do aspecto pessoal do pesquisador.

Este questionário foi dividido em duas partes: a primeira com o objetivo de verificar o tratamento dos dados na organização com uma visão mais abrangente, ou seja, tratamento/qualidade dos dados de modo geral e a segunda parte enfocando especificamente em projetos de DW e/ou DM.

Inicialmente foi enviado um email a um grupo composto por 60 profissionais de TI de organizações sediadas no Rio de Janeiro, convidando-os a participar da pesquisa. Muitos responderam que não se sentiam confortáveis em participar, pois nas organizações onde trabalham o tema qualidade de dados ainda é um assunto delicado de ser discutido; outros justificaram sua atitude pelo fato da empresa não possuir nenhum projeto, nem ferramenta de qualidade de dados e que as mesmas ainda são caras e também pelo fato de não existir um DW e/ou um DM e sim intenção de construí-lo. Porém houve colaboração no sentido de indicar outros profissionais envolvidos nesse tipo de projeto e que poderiam estar participando mais efetivamente da pesquisa.

3.4 Perfil dos Respondentes

Com relação ao perfil dos profissionais que responderam à pesquisa, totalizando 43, mais de 70% possuem experiência acima de dez anos na área de TI e cerca de 75% dos profissionais são prestadores de serviço da organização, porém quando se verifica o tempo em que atuam na área de DW e/ou DM, essa distribuição fica mais uniforme. (Figura 7).

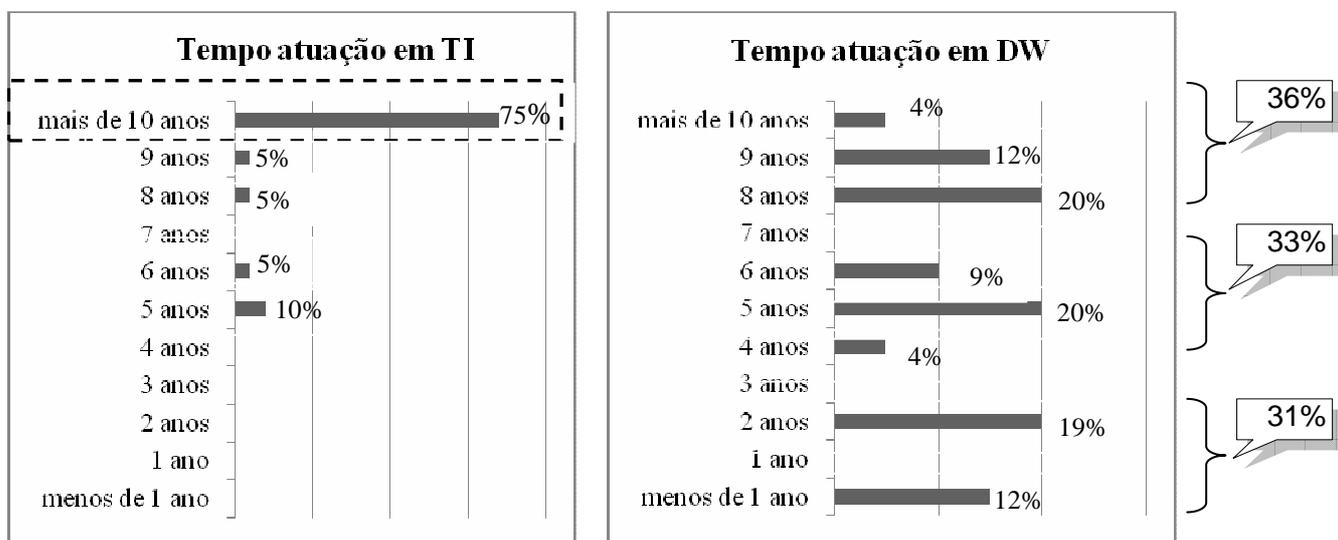


Figura 7 - Tempo de atuação em TI e em projetos de DW e/ou DM

Também foi observada na medição do tempo em que se deu o início do projeto de DW e/ou DM que 59% tiveram seu início há cinco anos ou menos. Já 64% tiveram a primeira aplicação OLAP disponibilizada aos usuários da organização em um tempo médio de no máximo cinco anos. (Figura 8)

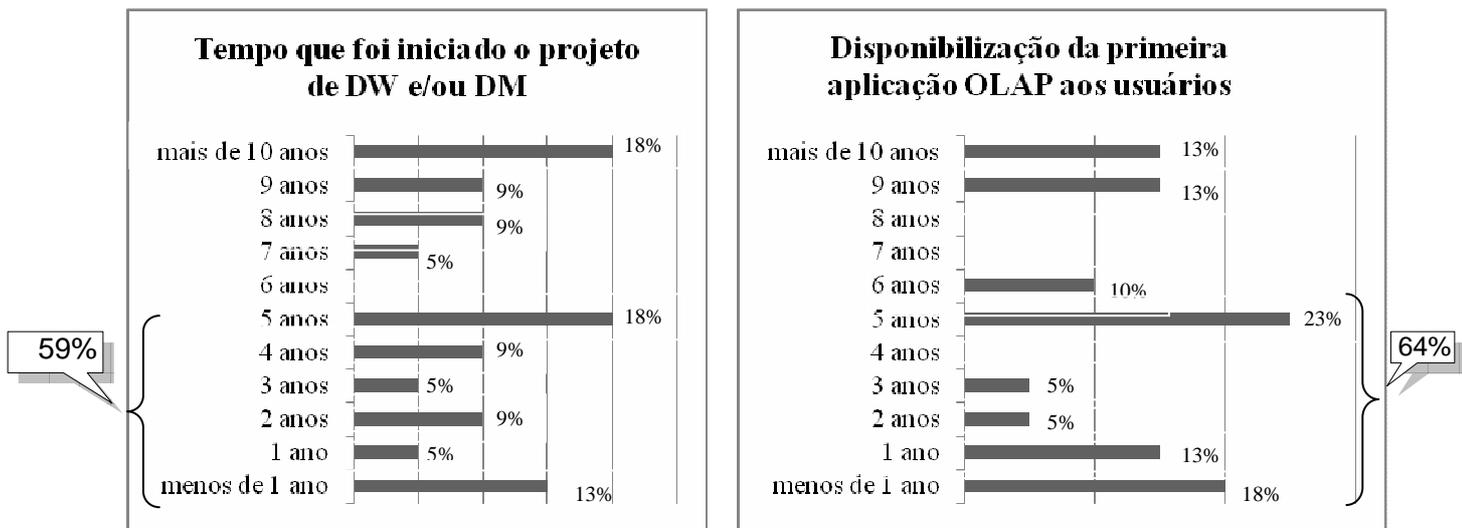


Figura 8 - Tempo de início projeto DW e/ou DM e disponibilização primeira aplicação OLAP

3.5 Resultados obtidos

Quanto às pessoas da administração da organização estarem envolvidas no patrocínio de projetos que implementem melhoria contínua de qualidade dos dados, observa-se que 37% identificam esse envolvimento e o restante afirma que este fato raramente acontece, nunca acontece ou não podem afirmar, indicando assim uma ausência de conscientização da importância da qualidade dos dados (Figura 9).

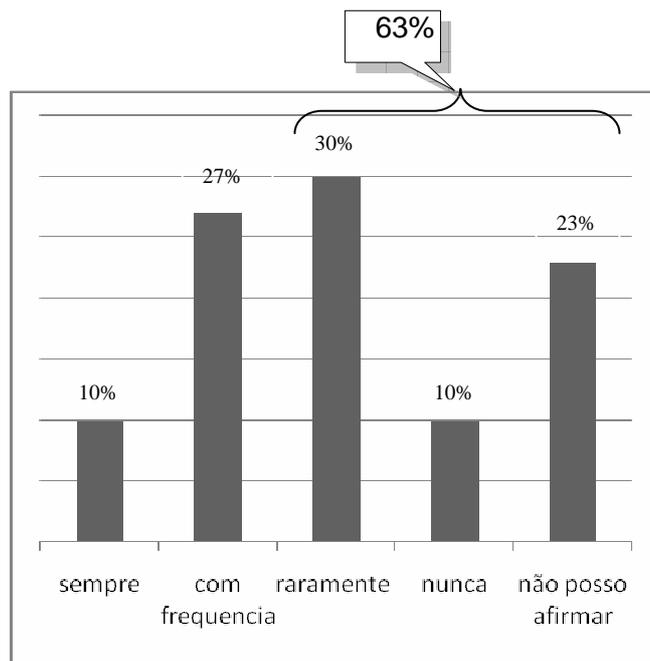


Figura 9 - Identificam pessoas na administração que patrocinam projetos de melhoria contínua qualidade dados

Porém observa-se que 50% dos respondentes conseguem identificar na organização coordenadores e/ou executores responsáveis por ações de melhoria de qualidade dos dados transacionais (Figura 10), mas apenas 42% dos respondentes conseguem identificar profissionais das áreas de negócio da organização envolvidos em ações de melhoria na qualidade dos dados, o que leva a crer que este assunto tem mais enfoque no âmbito da TI (Figura 11).

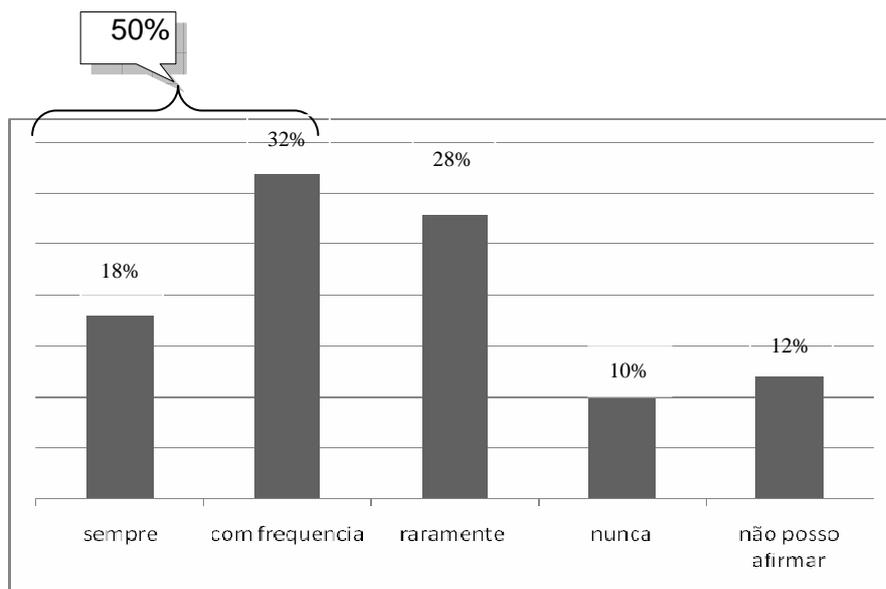


Figura 10 - Identificam pessoas responsáveis (coordenador/executor) por ações de melhoria na qualidade dados transacionais

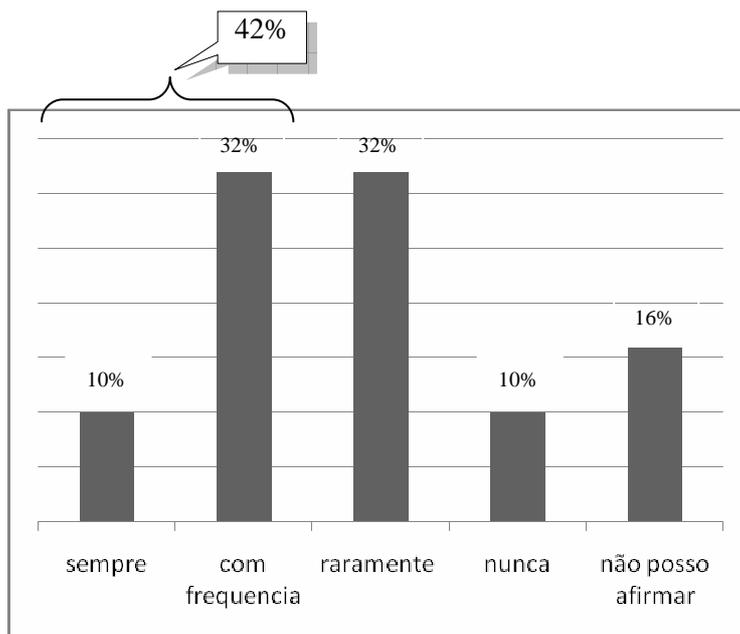


Figura 11 - Identificam pessoas pertencentes às áreas de negócio da organização envolvidas com ações de melhoria na qualidade dados

Cabe destacar a observação feita por um respondente registrando que as ações de melhoria na qualidade dos dados transacionais acontecem sempre de forma reativa, exemplificando com um tratamento efetuado em sua organização em relação aos endereços dos clientes, pelo fato de acontecer um retorno muito grande nas correspondências enviadas. Já outro respondente afirmou que a existência de uma coordenação de Métodos e Qualidade dentro da área usuária é fundamental para as pessoas das áreas de negócio da organização se envolverem em ações de melhoria na qualidade dos dados.

A maioria dos que responderam à pesquisa, ou seja, 95% reconhece que a utilização de dados com qualidade permitirá que a organização seja mais eficiente. Já 80% acham que ela se torna mais organizada e cerca de 70% percebem que ela se tornará mais ágil e competitiva, sugerindo assim um bom nível de conscientização do profissional de TI a cerca dos ganhos efetivos obtidos na organização. (Figura 12).

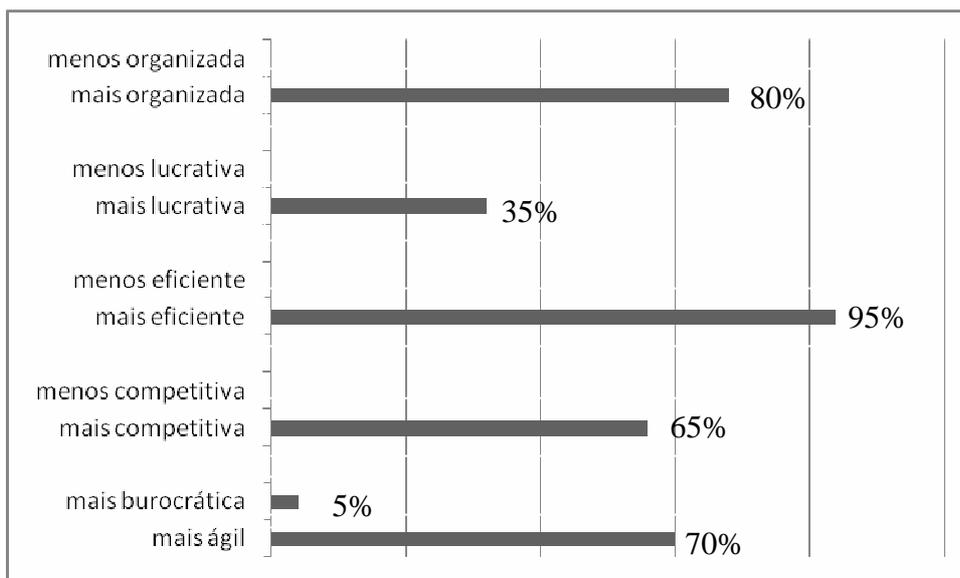


Figura 12 - Situação da organização ao utilizar dados com qualidade

No que se refere a projetos de DW e/ou DM, cerca de 65% dos pesquisados afirmam que a missão e os objetivos do DW e/ou DM encontram-se bem definidos (Figura 13).

Isto denota que apesar dessa base de dados ter como objetivo principal o de habilitar o trabalhador especializado, ou seja, o executivo, gerente e analista a tomar decisões melhores e mais rápidas (CHAUDHURI e DAYAL, 1997), persistem dúvidas, indefinições e desconhecimento à cerca dos efetivos objetivos deste projeto.

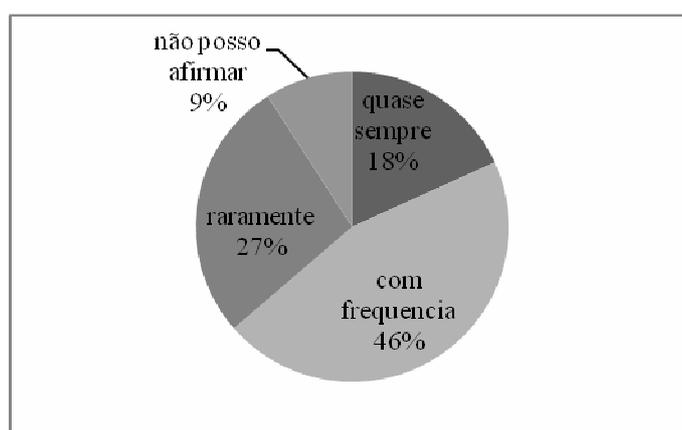


Figura 13 - Missão e objetivos do DW e/ou DM encontram-se bem definidos

Cerca de 70% afirmam que a documentação das fontes de dados é feita no início do projeto ou à medida que o projeto de DW e/ou DM vai necessitando (Figura 14). Isto sugere uma grande dependência da documentação dos dados em relação aos sistemas de apoio à decisão. Entretanto, como em geral tais iniciativas consideram um conjunto de

dados específico para a carga no DW/DMs, possivelmente nem todos os dados (ou menos as fontes de dados) estejam documentados. Esta falta de documentação pode levar a problemas relacionados à qualidade de dados independente dos esforços de integração dos dados.

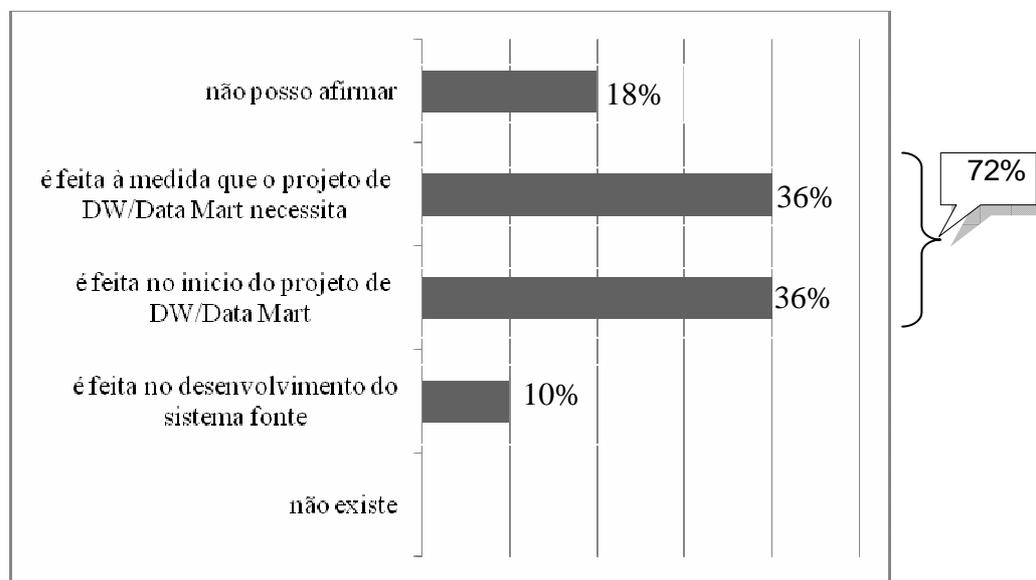


Figura 14 - Documentação das fontes de dados do projeto de DW e/ou DM

Ainda se observa que 50% dos pesquisados afirmam que a documentação dos modelos de dados é mantida pela equipe de TI, o que indica certo descaso com a documentação dos dados e, portanto, uma possível má gerência da qualidade de dados. 40% dos respondentes afirmam existir um processo que avalia a qualidade de dados existentes na organização, embora 50% dos pesquisados indiquem que o repositório de metadados, contendo o significado dos dados existentes atende parcialmente as expectativas dos usuários. Isto é curioso porque apesar da pouca documentação e processos de avaliação da qualidade dos dados praticamente inexistentes, os respondentes acreditam que as expectativas dos usuários quanto à qualidade dos dados estejam parcialmente atendidas.

Ainda cerca de 60% dos respondentes afirmam que existe (com frequência ou raramente) um repositório central de metadados, utilizado para documentar essas fontes de dados, provavelmente fornecidas pelas ferramentas utilizadas (Figura 15). O fato de 73% dos respondentes indicarem raramente, não posso afirmar ou nunca, indica um risco alto de não haver uma definição padronizada ou aceita do significado dos dados,

bem como sua origem, transformação e destino, ou mesmo quanto a seu acesso e manipulação.

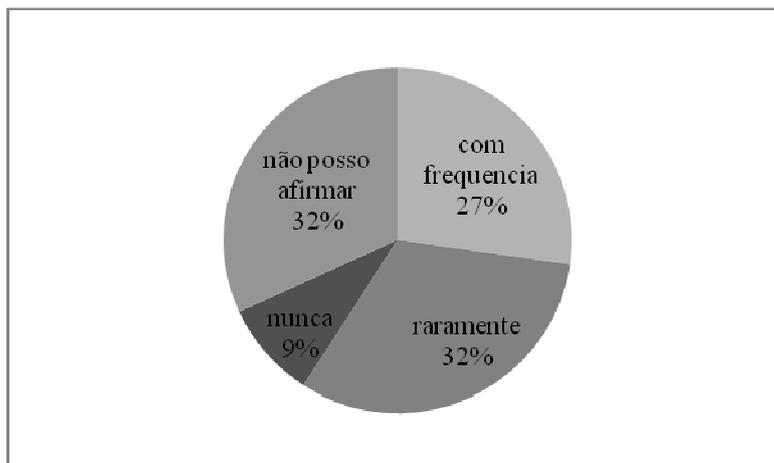


Figura 15 - Existência de um repositório central de metadados

Um dos respondentes afirmou que geralmente as ferramentas de ETL (Extração, Transformação e Carga) possuem um repositório central de metadados, e que ele considera ser suficiente para registrar toda documentação das fontes de dados utilizadas no DW. Já outro respondente acrescentou que somente o fato da existência de uma ferramenta e um repositório de metadados não é suficiente, uma vez que nem todos conhecem essa ferramenta e este repositório não se encontra integrado com o DW. Com isso não existe um compromisso de manter esse conteúdo atualizado. O mesmo acrescenta que recentemente estas informações foram publicadas no site da organização e que com isso ele acredita que o mesmo tenderá a ser mantido atualizado. Isto indica claramente a visão das empresas brasileiras de que a TI, em especial através dos processos de extração, transformação, carga e metadados, garante a qualidade de dados, independente da necessidade do usuário.

Existem diversas metodologias para a definição, medição, análise e melhoria da qualidade dos dados que podem ser encontradas na literatura. Na pergunta específica para verificar a utilização de alguma metodologia, destacou-se três: ISO 9001, GQM e TDQM (Descritos na seção 2.2.3). Porém mais de 80% afirmam que não utilizam uma metodologia que organiza conceitos e procedimentos para a definição, medição, análise e melhoria da qualidade dos dados (Figura 16).

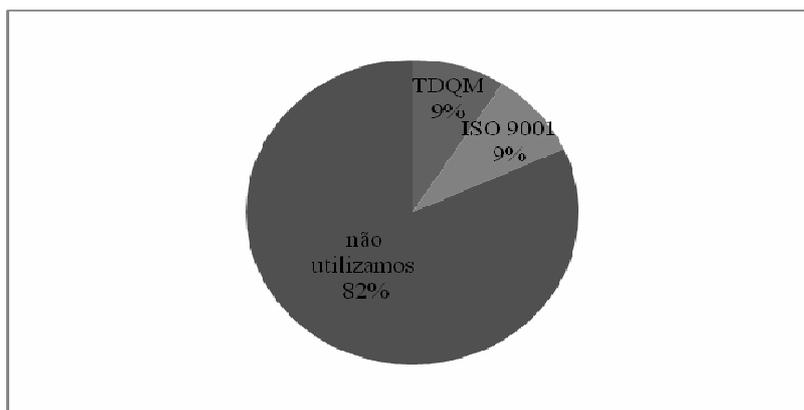


Figura 16 - Metodologia utilizada para definir conceitos e procedimentos para a definição, medição, análise e melhoria da qualidade dos dados

Em seguida foram feitas algumas perguntas que se referiam aos tipos de tratamento de dados efetuados nas fases do processo de ETL (extração, transformação e carga) e o quanto dos problemas encontrados eram devidamente tratados e registrados em um repositório de metadados (Figura 17). 28% dos respondentes indicam que conflitos de domínios ou necessidade de conversão não são tratados no processo de ETL (e outros 27% não podem afirmar). Além disto, 41% dos respondentes indicam que atributos inconsistentes não são facilmente detectados no processo de ETL (outros 18% não podem afirmar). Isto indica claramente problemas de qualidade de dados.

Atributos que possuem conflitos de domínios ou necessidade de conversão são tratados no processo de ETL:

Atributos inconsistentes são facilmente detectados no processo de ETL:

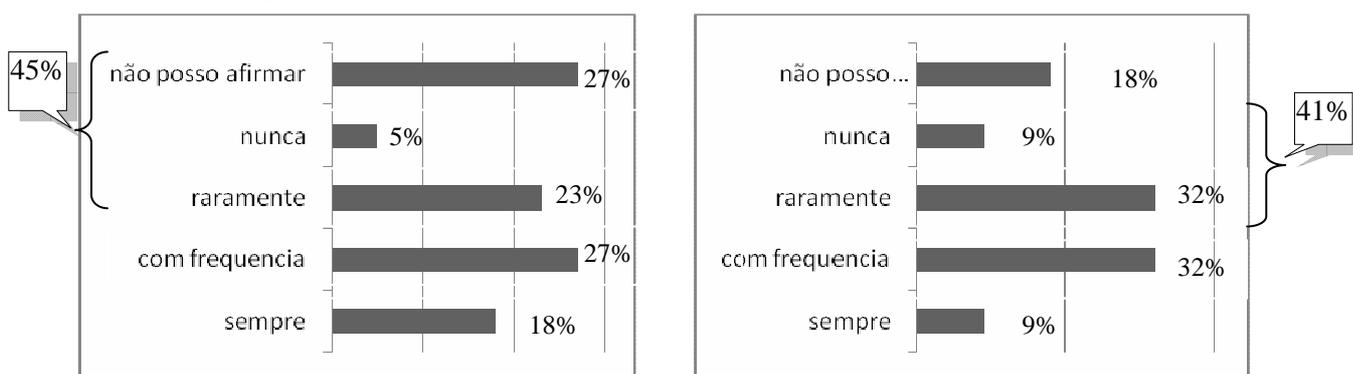
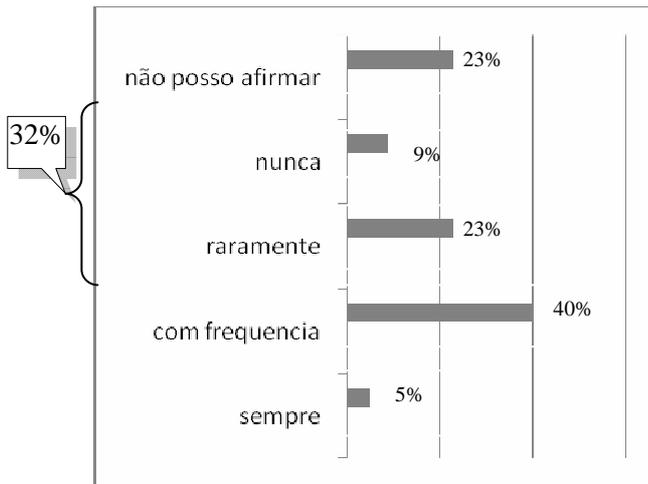


Figura 17 - Tratamento dos atributos no processo de ETL

Observa-se que em torno de 35% dos respondentes não conseguem identificar o tratamento e registro desses fatos, indicando assim os dados suspeitos de má qualidade, tais como: atributos inconsistentes, com conflitos de domínio ou necessidade de conversão, fontes de dados com regras de negócio inconsistentes e violações de regras de integridade nas fontes de dados. (Figura 18)

Fontes dados com regras negócio inconsistentes



Violação de regras de integridade nas fontes de dados

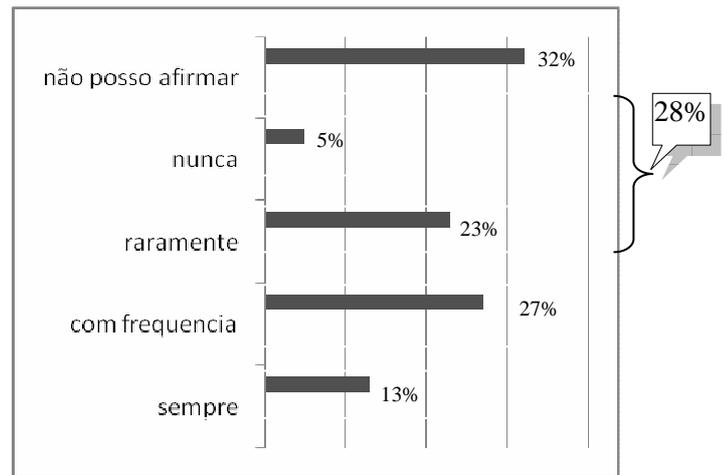


Figura 18 - Tratamento de regras de negócio no processo de ETL

Cerca de 65% dos pesquisados reconhecem que o processo de tratamento e limpeza dos dados deve gerar correção no sistema de origem para que posteriormente possam ser carregados no DW e/ou DM, porém vários afirmaram que na prática isto não acontece. (Figura 19). Cabe destacar que um dos respondentes afirmou que a identificação de inconsistência no processo de ETL é muito difícil em função da mudança da regra de negócio, e acrescenta que com isso o próprio usuário do BI comunica à equipe quando identifica um problema.

Todo e qualquer processo de tratamento e limpeza dos dados deve acontecer antes da inclusão no DW e/ou Data

O processo de tratamento e limpeza dos dados deve gerar correção no sistema de origem dos dados para que posterior-mente possa ser carregado no DW e/ou Data Mart

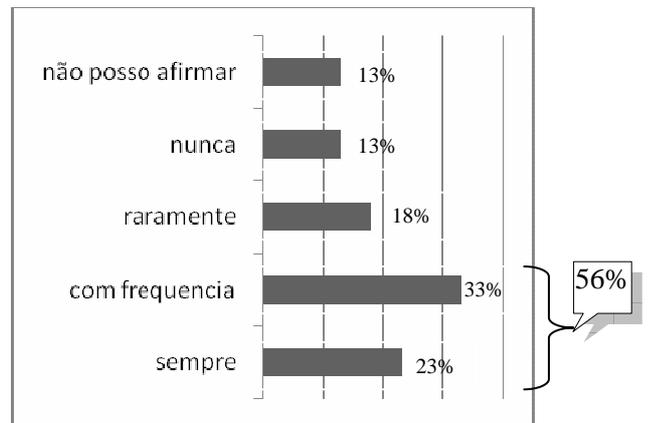
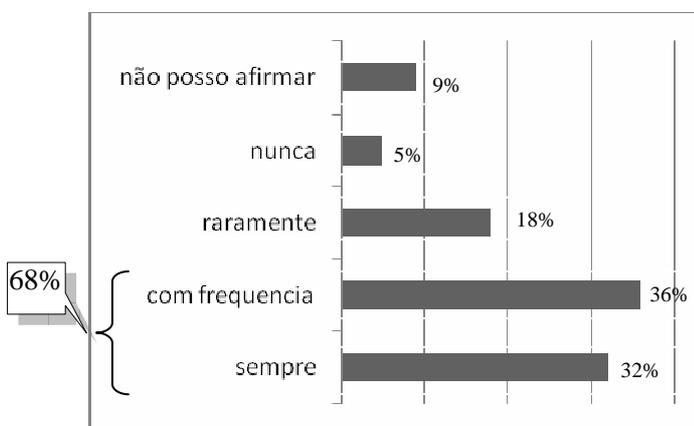


Figura 19 - Avaliação tratamento de dados efetuado no processo de ETL

Em torno de 50% dos respondentes afirmam que regras de sumarização dos dados encontram-se identificadas e mapeadas em um repositório de metadados e 40% afirmam que na sumarização de um determinado dado, o grau no qual este conteúdo é válido, é quantificado e armazenado em um repositório de metadados. (Figura 20)

Na agregação de valores, a regra de sumarização dos dados envolvidos encontra-se identificada e mapeada em um repositório de metadados

Na sumarização de um determinado dado, o grau no qual este conteúdo é válido, é quantificado e armazenado

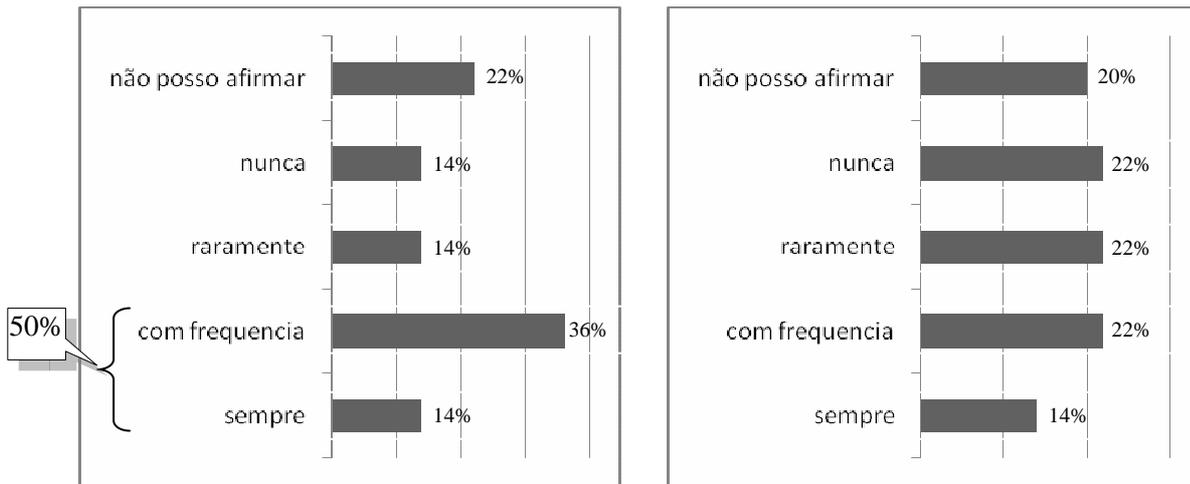


Figura 20 - Avaliação tratamento de dados na sumarização

Mais de 70% dos respondentes acreditam que o processo de tratamento e limpeza dos dados garante cerca de 80 a 99% da qualidade de dados no DW e/ou DM (Figura 21).

Apesar de a pesquisa mostrar que existe uma minoria que possui a crença de que se pode obter 100% de garantida da qualidade dos dados, um respondente complementou que nunca se poderá garantir 100% em função da velocidade da mudança do negócio.

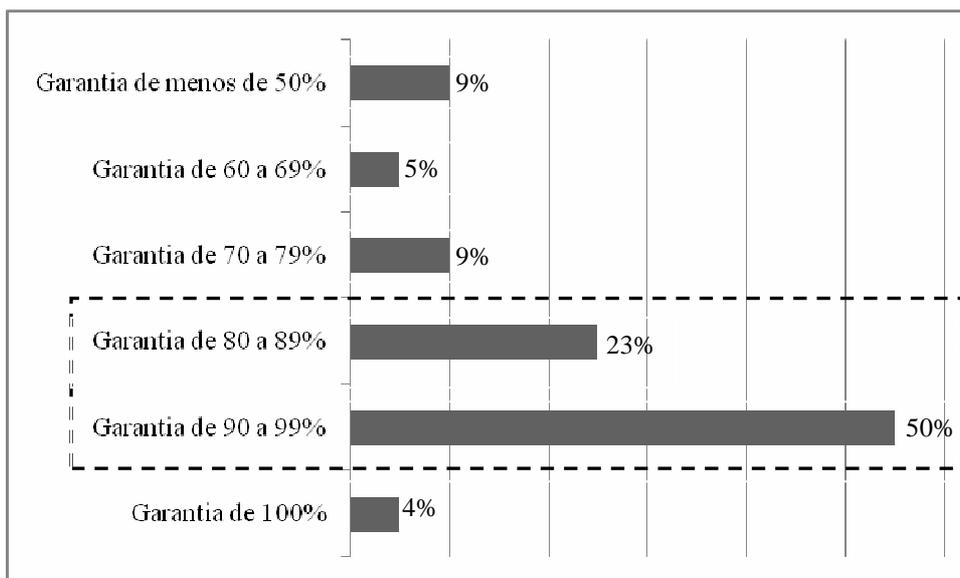


Figura 21 - Avaliação da garantia da qualidade de dados no DW e/ou Data Mart executada pelo processo de tratamento e limpeza dos dados

Formal ou informalmente, os profissionais de TI se mantêm atualizados sobre a satisfação dos usuários do DW e/ou DM, porém 60% afirmam que os usuários continuam manifestando a necessidade de ter mais informações sobre a qualidade do conteúdo do mesmo. (Figura 22) E para finalizar, cerca de 60% dos pesquisados consideram que mapear critérios de qualidade de dados no DW e/ou DM é uma tarefa extremamente árdua pela ausência de metodologias ou ferramentas que os apoiem neste processo.

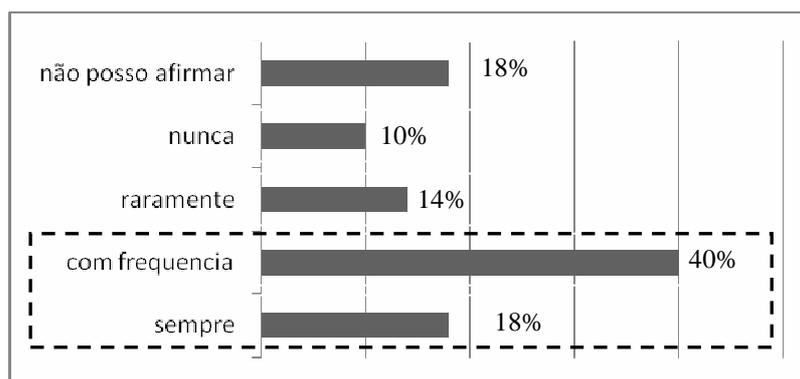


Figura 22 - Quando um usuário manifesta necessidade de mais informações sobre a qualidade do conteúdo

Um respondente complementou a questão informando que os usuários procuram sempre saber quanto às regras que estão sendo utilizadas no DW e acrescenta também que existem ferramentas para auxiliar o processo de mapeamento de critérios de qualidade de dados no DW e/ou DM, porém ainda muito custosas e não tão precisas.

3.6 Conclusão e comentários finais

Existe uma dificuldade por parte das equipes de TI tanto no que se refere ao entendimento de conceitos quanto à explicitação de critérios de qualidade de dados. Para tal explicitação, acreditava-se que uma sistemática pudesse vir a apoiar a equipe de TI envolvida em projetos de DW e/ou DM, sendo uma contribuição significativa para auxiliar e agilizar o processo de tratamento de limpeza e explicitação da qualidade dos dados.

Porém, após pesquisa bibliográfica, análise do resultado da pesquisa aplicada, bem como discussões informais com profissionais experientes em projetos de BI, inclusive levantando-se a relação do alto índice de insucesso desses projetos estarem diretamente ligados à questão da qualidade de dados, ficou patente que, de fato, esse assunto transcende a área de TI, devendo ser tratado por toda organização.

Ainda se percebe que o assunto qualidade de dados é tratado nas organizações com certo tabu e que sua importância no contexto organizacional não está totalmente assimilada. Por outro lado, verifica-se uma crença de que a resolução de quase todos os problemas de qualidade de dados está no processo de extração, transformação e carga do DW, por parte do pessoal de TI (principalmente os envolvidos em projetos de BI).

Em função destes equívocos é que o tratamento da informação deve ser reconhecido e tratado como qualquer outro recurso tangível existente na organização. Essencialmente, a pesquisa evidenciou a necessidade da inserção de projetos relacionados à qualidade de dados dentro de um programa de Governança de Dados.

4. DATAGOV-ADM: Uma proposta de Governança de Dados baseada no TOGAF-ADM

O ativo mais importante de qualquer empresa atualmente são os dados que ela dispõe, mantém e disponibiliza para seus fornecedores, clientes, parceiros, etc. As organizações, em função da crescente competição no mercado, de exigências legais mais rigorosas, ou por demandas de transparência de gestão, se tornaram cada vez mais dependentes dos dados. Porém, estes continuam com problemas: inconsistência, duplicidade, informações incompletas e disponibilização fora do prazo são apenas alguns dos obstáculos mais comuns encontrados em praticamente todas as organizações.

A Governança de Dados corporativos está relacionada à forma de tratar o dado como um recurso tangível da organização. Isso envolve processos, políticas, padronização, organização e tecnologias necessárias para manipular e assegurar a disponibilidade, acessibilidade, qualidade, consistência, auditabilidade e segurança dos dados na empresa (FISHER, 2009).

Um método de Governança de Dados pode ajudar as organizações a manipularem os dados de forma mais eficiente, devendo prover definições consistentes, estabelecer uma administração de dados na organização, mensurar e rastrear a qualidade dos dados transacionais e analíticos utilizados ao longo da organização.

A Arquitetura Empresarial, por ser um grande modelo que documenta e representa todos os comportamentos que fazem uma organização funcionar de forma integrada e sincronizada, pode servir de base para uma metodologia de Governança de Dados que possibilite o alinhamento da tecnologia da informação e o negócio. O método ADM descrito no TOGAF apresenta uma sequência iterativa de passos para criação e desenvolvimento de uma arquitetura empresarial completa. Com base no método ADM do TOGAF, o objetivo deste trabalho é propor uma sequência iterativa estruturada, capaz de elucidar todos os passos necessários para uma efetiva Governança de Dados em uma organização (Figura 23).

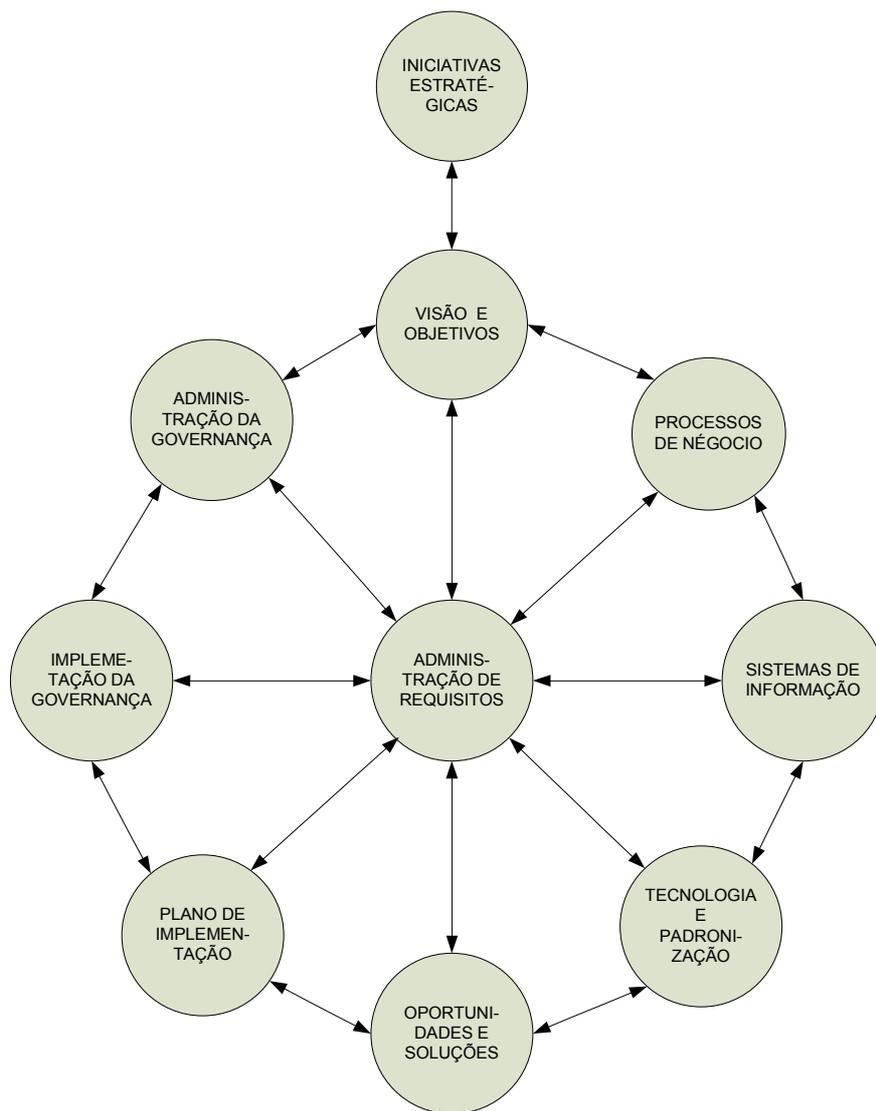


Figura 23 – DATAGOV-ADM: Governança Dados baseada no TOGAF-ADM

Nas seções subseqüentes se descreve cada etapa desta proposta, detalhando os passos e assuntos que devem ser tratados.

4.1 Iniciativas estratégicas

As organizações necessitam de informações oportunas e conhecimentos personalizados para efetivamente auxiliar os seus processos decisórios e a sua gestão empresarial, principalmente por estarem enfrentando um mercado altamente competitivo e globalizado. De forma permanente, elas buscam adequação e ajustes entre suas funções e operações cotidianas com as reais e efetivas necessidades do meio ambiente interno e

externo onde se encontram inseridas, por meio de decisões e ações de seus gestores (FISHER, 2009).

O planejamento estratégico é um processo de determinação dos principais objetivos de uma organização, das políticas e estratégias que a governarão, do uso e disponibilização dos recursos para a realização dos objetivos. Esse processo é composto por premissas, planejamento propriamente dito, implementação e revisão. Ele também pode ser chamado de Plano de Negócios (BHALLA, 1987; ABRAMS, 1991). Seus objetivos principais são definir um posicionamento estratégico que crie e fortaleça a vantagem competitiva da empresa, e estabelecer ações estratégicas para tornar este posicionamento realidade. Esse processo pode ser visto como uma atividade gerencial que permite aos executivos estabelecerem um rumo para a organização, em que se busca certo nível de otimização no relacionamento entre organização e ambiente, e que corresponde ao estabelecimento de um conjunto de providências a serem tomadas pelos gestores para um futuro que tende a ser, cada vez mais, diferente do passado (OLIVEIRA, 1999).

O olhar estratégico faz com que um gestor passe a gerenciar melhor os problemas do dia-a-dia porque o seu foco muda da “discussão dos problemas” para um questionamento sobre “como transformar os problemas que surgiram em oportunidades para a empresa”. Para maior eficiência operacional e gerencial a estratégia de qualidade de dados deve caminhar junto com a estratégia do negócio. Isto inclui objetivos estratégicos que são perseguidos pela equipe de Governança de Dados e como eles encontram-se alinhados com a estratégia de negócio da empresa e seu escopo funcional global. A qualidade de dados é de responsabilidade de todos os colaboradores na organização que produzem, inserem, extraem ou analisam dados a partir de sistemas de informação, eletrônicos ou manuais, dentro de seus perfis de autorização. Informações necessitam encontrar-se em um nível suficiente de acurácia, perfeição e estarem no tempo certo (*timing*) para que os gestores possam utilizá-las para acompanhamento das operações, monitoramento e apoio à tomada de decisão. Por isso, a inclusão e conjugação da Governança de Dados ao planejamento estratégico organizacional devem acontecer de forma contínua e natural, possibilitando o seu tratamento a nível organizacional e criando mecanismos que incentivem a prática de administração da qualidade dos dados em toda a organização.

Logo, a fase preliminar deste framework sugere que a organização contemple no Planejamento Estratégico a definição das estratégias de qualidade de dados, pois:

1. Medidas de desempenho e administração são fundamentais para o sucesso da organização;
2. a efetiva utilização dessas medidas depende diretamente dos dados precisos e confiáveis;
3. a boa qualidade da informação é essencial para apoiar o processo decisório em qualquer nível organizacional;
4. é vital que a informação utilizada para informar, administrar e planejar atividades seja baseada em dados de boa qualidade.

A Figura 24 representa graficamente a utilização da informação para melhor tomada de decisão:

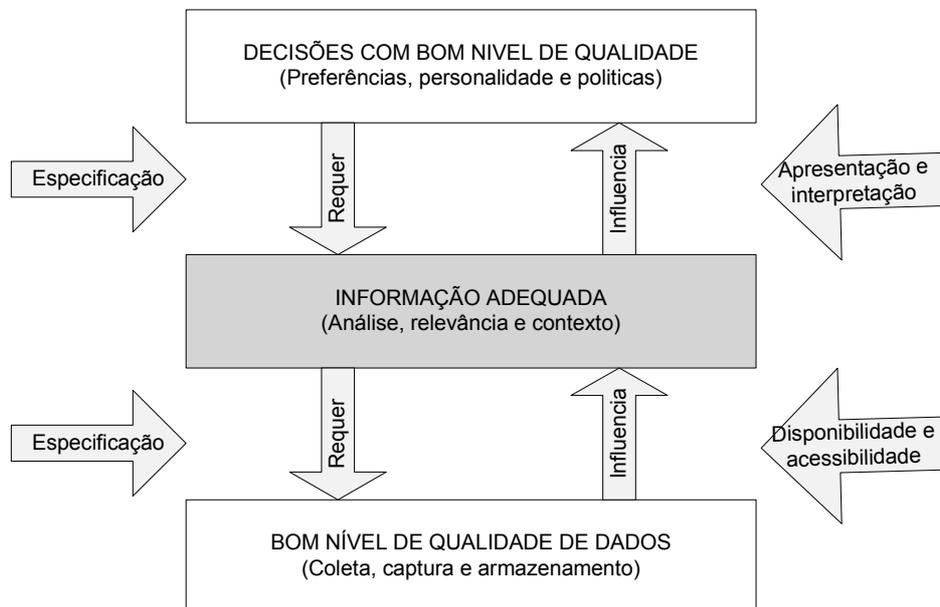


Figura 24 - Fases na produção e utilização da informação (MCGEE e PRUSAK, 1994)

A utilização estratégica da informação para o negócio revela duas possibilidades reais, de acordo com MCGEE e PRUSAK (1994): a identificação de alternativas inovadoras e a identificação de situações que poderão interferir na própria sobrevivência da empresa. Ainda de acordo com os autores, a definição da Estratégia da Informação considera os elementos de ‘Posicionamento e Escopo’, ‘Governança’, e ‘Competências Específicas da Empresa’. Esses elementos são de vital importância, não apenas na esfera da organização, mas também para avaliar e decidir sobre interações e operações com o mercado. A estratégia da informação deve indicar um direcionamento para a empresa, trazendo à tona o conhecimento de quais informações são importantes e o domínio do

momento adequado de seu uso para a tomada de decisão.

4.1.1 Definição de papéis e responsabilidades

Cabe à alta direção da organização determinar as responsabilidades para implantação da Governança de Dados e para tal deve definir:

1. *Comitê de política e gestão de dados*, formada por gestores das áreas de negócio e de profissionais de TI, com as seguintes atribuições principais:
 - Manter a relação de sistemas e coleção de dados;
 - coordenar a revisão de todos os sistemas de informação;
 - manter uma lista de recomendações a serem tomadas para tornar os sistemas mais eficientes no tratamento de captura dos dados;
 - verificar se todas as recomendações foram implementadas e se os dados existentes foram devidamente revistos e corrigidos;
 - divulgar os princípios de qualidade de dados, ou seja, regras e ações que devem ser adotadas em toda organização;
 - revisar desempenho e desafios.
2. *Auditoria interna*, com o foco em auditoria em qualidade de dados:
 - Recomendar melhorias nos sistemas de informação à luz da avaliação da qualidade dos dados;
 - incorporar assuntos de qualidade de dados nas rotinas de auditoria;
 - fornecer assistência, orientação e apontar necessidade de treinamento, onde e quando for necessário.
3. *Avaliação de riscos e conformidade*, que devem:
 - a. Avaliar se a iniciativa aumentou a habilidade da organização em identificar e gerenciar riscos e requerimentos legais, regulatórios e contratuais;
 - b. Registrar e fazer recomendações necessárias de melhorias.
4. *Stakeholders de dados*, gestores das áreas de negócio, com o foco em orientações de qualidade dados, com as seguintes atribuições:
 - Definir e comunicar as diretrizes da governança na organização;
 - comunicar a implementação de novas políticas à organização;
 - planejar e definir requisitos de informação;
 - definir e assegurar controles da informação.

São necessárias qualificações às pessoas integrantes destes papéis, que devem ter um bom conhecimento à cerca do negócio e total entendimento dos dados que são requeridos. Também devem ter um bom nível de conhecimento das informações, processos e regras. Outro quesito importante é o conhecimento organizacional e cultural e as perspectivas da indústria onde está inserido o negócio da empresa.

Tais pessoas devem saber se comunicar, tanto por escrito como verbalmente, com clareza e precisão. Além disto, devem saber trabalhar em equipes sem gerar conflitos, e é importante que procurem ser adaptáveis, objetivas, e saibam agir com diplomacia. Não obstante, devem estar sempre ponderando as necessidades funcionais e locais com as com as necessidades da organização.

4.1.2 Definição de políticas e procedimentos

A Governança de Dados deve orquestrar pessoas, processos e procedimentos para criar uma visão empresarial consistente, com o objetivo de aumentar a confiança nos processos decisórios, diminuir os riscos de não conformidade no que se refere às regulamentações e aumentar a segurança dos dados.

Assim, a governança deve criar e implantar o que for considerado como dados com qualidade aceitáveis, através da utilização de políticas de negócios que guiem a coleção e administração dos dados. Deve-se buscar a confecção de um manual onde constem as políticas e procedimentos esperados na condução da Governança de Dados, servindo como ponto de referência para todas as pessoas envolvidas. Devem ser definidas políticas e procedimentos para padronização de dados e também para modelagem de dados. Deve-se definir padrões de registro, por exemplo, se for adotar o critério de rastreabilidade, utilizar um padrão para registrar quem e quando foi feita a última atualização, adotando em todas as estruturas do modelo dados, bem como nas aplicações. E por último deve-se definir a estratégia de comunicação e frequência com que as atualizações, métricas e realizações do programa de governança serão divulgadas.

4.2 Visão e Objetivos

Esta é a etapa inicial de interação do framework proposto, determinando escopo, restrições, expectativas e validando o contexto da governança. Esta abordagem tem como ponto de partida a identificação dos pilares da organização:

Visão é o conceito que abrange as aspirações futuras, ou seja, o ponto onde a organização quer alcançar: Como sua organização estará no futuro utilizando um programa de Governança de Dados? E como estará se não tiver?

Objetivos diz respeito à definição de metas, propósitos e resultados a serem alcançados no final. Definir também acordos sobre o escopo do trabalho, ou seja, determinando escopo, restrições e expectativas e métodos a serem adotados.

O estabelecimento destes pilares serve como base e referência para todo desdobramento das ações relacionadas à Governança de Dados da instituição, desde os objetivos e diretrizes estratégicos, e seu desdobramento em planos de ação, em indicadores e em metas de desempenho, até a implementação e o contínuo processo de avaliação e monitoramento do desempenho da instituição. Para a análise dos pilares, serão levadas em consideração as seguintes questões críticas:

1. Estão definidos e divulgados adequadamente?
2. Fazem parte da cultura da instituição?
3. São referências para a atuação dos indivíduos da instituição?

A tradução destes objetivos e diretrizes nas ações, indicadores e metas de desempenho relacionadas à Governança de Dados (como parte integrante de análise de todo o modelo da gestão estratégica, bem como sua implementação também) devem ser avaliados.

4.2.1 Definições e Fundamentação estratégica

É necessário acrescentar definições que descreverão de uma forma mais clara e objetiva o escopo da Governança de Dados, o resultado que se quer obter e como serão medidas as implementações efetuadas, explicitando assim a eficácia da governança. Para tal devem ser consideradas as seguintes questões:

1. Que áreas de negócio serão contempladas?
2. Que repositório específico será utilizado?

3. Qual a situação atual que já é percebida por quem o utiliza?
4. Como eles deverão ficar após a Governança de Dados?
5. Que processos e práticas da administração de dados serão afetados e como?
6. Que processos e práticas de negócio e tecnologia serão afetados e como?
7. Como as pessoas trabalharão juntas da forma que não se encontram agora, e qual é o resultado esperado?

4.2.2 Regulamentações e *Compliance*

O gerenciamento de riscos nas organizações visa à redução dos impactos dos riscos, caso eles venham a se materializar. Este gerenciamento vai além da avaliação de suas probabilidades de perda, abordando também o estabelecimento de medidas que visem à mitigação dos riscos ao longo do tempo. Para gerenciar os riscos, são estabelecidos controles que minimizam seus impactos, aprimorando o sistema de controles internos da instituição.

Todas as empresas são impactadas por regulações governamentais, financeiras e específicas de sua área de atuação. Várias dessas regulamentações determinam como dados e informações devem ser manuseados. Uma boa razão para implementar Governança de Dados é cumprir determinações oriundas dessas regulações. A Governança de Dados deve orientar a implementação de controles adequados para assegurar, documentar e monitorar o cumprimento dessas regras no que diz respeito às regulações de dados.

A Lei Sarbanes-Oxley (PROFORMA, 2003), conhecida também como SOX, é uma lei cujo objetivo é justamente aperfeiçoar os controles financeiros das empresas e apresentar eficiência na governança corporativa, a fim de evitar que aconteçam escândalos e prejuízos financeiros. Esta lei visa garantir a transparência na gestão financeira das organizações, credibilidade na contabilidade, auditoria e a segurança das informações para que sejam realmente confiáveis, evitando assim fraudes, fuga de investidores, etc. A SOX despertou nos executivos um interesse urgente de saber o quanto os seus dados encontram-se consistentes, com acurácia, e atualizados. Existem várias outras regulamentações com significativa influência em como os recursos-informação são tratados, manuseados e administrados na organização, tais como:

1. HIPPA (The Health Information Protection and Portability Act): é uma lei federal americana que desde 1996 exige que empregados, médicos e seguradoras tratem com privacidade e segurança informações sobre a saúde de um paciente.
2. Basel II New Accord: Desde 2006, instituições financeiras que desejam operar em países pertencentes à União Européia, são obrigados a reportar informações que demonstrem liquidez.
3. Solvency II: Regulamentações para as seguradoras que atuam na União Européia
4. PCI-DSS: Padrões de segurança de dados na indústria de cartão de crédito.
5. GASB (Government Accounting Standards Board) e FASB (Financial Accounting Standards Board) padrões contábeis com implicações em como o recurso-informação tem que ser tratado.

Poderão existir outras regulamentações, além das citadas acima, que são ditadas pelos Órgãos Reguladores das atividades econômicas, tais como, Instituições Financeiras (BACEN), Seguro (SUSEP), Previdência (PREVIC), Telecomunicações (ANATEL), Transporte (ANTT), etc. que ditam regras e regulamentam a área de negócio na qual a empresa está inserida. A Governança de Dados deverá identificar quais regulamentações devem ser atendidas e responder da melhor forma possível as seguintes questões relativas ao cumprimento de leis regulatórias:

1. Quão relevante é a regulamentação? Porque é tão importante esta regulamentação para a empresa?
2. Como interpretamos isto? Que políticas e procedimentos são necessários?
3. Estamos cumprindo hoje? Como o estamos fazendo?
4. Como deveremos cumprir no futuro? O que deixaremos de ganhar caso deixemos de cumprir? Quando estaremos cumprindo?
5. Como demonstramos e provamos que estamos cumprindo as regras e leis regulatórias?
6. Como monitoramos o cumprimento das regras e leis? Revemos com uma periodicidade definida?
7. Como identificamos e reportamos o não cumprimento de alguma norma?
8. Como administramos e retificamos o não cumprimento de alguma norma?

4.2.3 Gestão de Risco Operacional e a Tecnologia da informação.

Um conjunto de ações deve ser adotado com o objetivo de mitigar os riscos operacionais relacionados à segurança de Tecnologia da Informação. Entre elas, estão aquelas dispostas pelo Disaster Recovery Institute (DRI), órgão mundial de padronização de planos de continuidade de negócios (que se utiliza das melhores práticas de mercado) e ações relacionadas em documentos internos da organização, elaborados a partir de referências baseadas como, por exemplo, em normas nacionais. Utilizando-se da metodologia de análise de inspeção física, a empresa deve buscar identificar e analisar os riscos e as vulnerabilidades passíveis de incidentes de segurança, visando a ações de melhoria que contribuam com a segurança da informação e com a execução/criação de planos de contingência. Um exemplo do resultado decorrente desse procedimento é a certificação pela ABNT do local onde se encontram os servidores e demais equipamentos que dão suporte aos dados críticos da empresa.

Também se faz necessária a definição de um plano de contingência e continuidade de negócios. Este deve ser disponibilizado por meio de comunicação interna, onde seja possível obter informações sobre o cronograma dos testes dos planos de contingência e continuidade de negócios.

A empresa deve manter o seu plano de continuidade de negócios estruturado por metodologia voltada aos seguintes focos: Administração de Crise; Resposta a Incidentes; Contingência Operacional, Recuperação de Desastre, e Testes e Simulações.

Adicionalmente, a área responsável pela Segurança da Informação contribui para o direcionamento das ações voltadas à continuidade de negócios dos processos críticos de TI da empresa, identificados preliminarmente com a realização da Análise de Impacto dos Negócios (*Business Impact Analysis*). Uma área específica da empresa deve assegurar a continuidade de negócios da rede de telecomunicações, realizando o monitoramento constante das linhas de comunicações de dados, acionando imediatamente a contingência, em caso da detecção de algum tipo de problema.

4.3 Processos de Negócio

Na segunda etapa deste framework, os processos de negócio devem ser definidos ou reavaliados buscando-se a realização dos objetivos identificados em um Modelo de Objetivos do Negócio. Levantamentos feitos com os envolvidos no negócio também

devem ser realizados para fornecer subsídios à definição dos processos de negócio. Devem-se identificar os principais processos de negócio, suas relações com os recursos-informação (entradas, saídas, fornecedores, controles e objetivo) e a sequência de execução dos mesmos. O mais importante é modelar o comportamento dos recursos-informação, nos casos em que ocorram várias alterações ao longo dos processos de negócio, já que esta dinâmica de alterações precisa ser mais bem entendida.

A construção da visão geral da arquitetura de processos é feita com base em um processo colaborativo no qual se listam as principais classes de processos da organização (ou funções da organização e suas definições): vendas, marketing, produção, etc. Para isso utiliza-se o conceito de classificação e evita-se redundância nas descrições.

Os processos sob o enfoque de gerenciamento da informação definem como a empresa obtém, distribui e utiliza a informação. DAVENPORT (2001) sugere a divisão do gerenciamento em quatro etapas: (a) determinação das necessidades; (b) obtenção das informações; (c) distribuição; e (d) uso da informação. Para determinação das necessidades, deve-se analisar o que se precisa e cruzar com o que se tem. As informações obtidas devem obedecer a padrões de classificação, formatação e estruturação. No gerenciamento da distribuição, é necessário que a análise da estrutura política, hierárquica e física seja analisada. E no que tange ao uso da informação, deve-se promover a identificação e acompanhamento dos usuários finais.

4.4 Sistemas de Informação

A definição de um modelo de arquitetura da informação da organização facilita a criação, o uso e o compartilhamento otimizados da informação pelo negócio para manter a integridade e ser flexível, funcional, com boa relação custo-benefício, rápido, seguro e mais resistente a falhas.

Por isso, nesta etapa, os sistemas de informação devem ser revistos, avaliando-os sob o enfoque da forma como está sendo tratado o recurso-informação. A construção da visão geral da arquitetura de informação permite a implementação de sistemas com mais qualidade e administração de seus dados. O resultado esperado é a definição de um conjunto de sistemas de informação e o nível de acoplamento entre eles. A perspectiva dos sistemas de informação combina o desenvolvimento de um modelo de dados corporativo a nível lógico, o desenho da arquitetura de dados, que lida, entre outras

coisas, com a validade dos objetos de dados corporativos no contexto organizacional, e a definição de um sistema de suporte.

As principais atividades a serem realizadas nesta parte da metodologia estão apresentadas abaixo:

1. Identificar objetos de dados corporativos;
2. Desenvolver definições únicas e não ambíguas para objetos de dados corporativos;
3. Estabelecer um dicionário de dados e metadados de negócio onde deverá constar o mapeamento de onde os dados são criados, mantidos, propagados e entregues atendendo ao objetivo do negócio, bem como as áreas que o utilizam (tanto a nível operacional, tático e estratégico);
4. Identificar dimensões de qualidade para objetos de dados corporativos;
5. Definir o armazenamento dos dados;
6. Estabelecer uma infra-estrutura corporativa de dados;
7. Definir princípios para distribuição de dados corporativos entre sistemas;
8. Definir a utilização de ferramentas e métodos de limpeza de dados.

Também é fundamental manter um dicionário de dados corporativo que incorpore as regras de sintaxe dos dados da organização. Este dicionário permite o compartilhamento dos elementos de dados entre os aplicativos e sistemas, promovendo um entendimento comum em toda a organização e prevenindo a criação de elementos de dados incompatíveis.

Os principais processos empresariais identificados durante o planejamento da arquitetura da informação vão indicar as propriedades com maior possibilidade de serem gerenciadas. Então, os dados corporativos devem ser disponibilizados para pesquisa e análise do que os usuários podem encontrar, a forma como as informações serão apresentadas e qual será o impacto no rastreamento. É possível que na maior parte das vezes, não se queiram exibir todos os dados, mas somente aqueles referentes a determinadas atividades. Por exemplo, uma organização com vários escritórios em varias regiões talvez queira exibir os resultados somente para escritórios específicos em algumas regiões-chaves, dependendo das necessidades da empresa. Logo, deve-se decidir quais tipos de dados corporativos e suas respectivas propriedades que serão incluídos.

4.5 Tecnologia e padronização

A estrutura de apoio para o ambiente informacional deve ser construída por peritos em tecnologia, que agreguem valor à informação, nesta quarta etapa. Devem definir e implementar procedimentos que assegurem a integridade de todos os dados armazenados na forma eletrônica, tais como banco de dados, *data warehouses* e arquivos de dados. Para tal faz-se necessário: fornecer apoio a infraestrutura de banco de dados, implementando e controlando banco de dados; definir regras de utilização de fontes de dados externas; planejar a recuperação de dados; estabelecer níveis de serviços para o desempenho do banco de dados; implementar regras para arquivar, reter e eliminar dados e apoiar bancos de dados “especialistas”.

Os peritos em tecnologia devem também definir e estabelecer políticas e procedimentos para identificar e aplicar requisitos de segurança aplicáveis ao recebimento, processamento, armazenamento físico e saída de dados para atender aos objetivos de negócio, à política de segurança da organização e a exigências regulatórias.

A subseções seguintes apresentam as diferentes perspectivas do processo de Governança de Dados que devem ser apoiados pela tecnologia de informação.

4.5.1 Qualidade de Dados

Como já abordado no Capítulo 2, qualidade de dados é um conceito complexo, com significados diversos para diferentes pessoas. De acordo com o que foi comentado anteriormente, pode-se apontar a qualidade dos dados como sendo uma medida de concordância entre as visões do dado, apresentada pelos sistemas de informação e o mesmo dado no mundo real. Para tal faz-se necessária, inicialmente, a definição de requisitos de qualidade que nortearão a governança dos dados.

4.5.1.1 Definição de requisitos de qualidade de dados

Aqui cabe definir que requisitos são importantes e primordiais de serem mensurados e acompanhados na organização. As dimensões de qualidade, com as respectivas definições para futura utilização, foram detalhadas no capítulo 2.

4.5.1.2 Avaliação dos Dados

Este é o processo de examinar os dados disponíveis em uma base de dados ou arquivo e coletar informações e estatísticas sobre esses dados. O objetivo dessa estatística é

mostrar que os dados existentes podem ser utilizados em outros propósitos com confiança e segurança. Também deve proporcionar métricas de qualidade, incluindo se os dados estão de acordo com padrões estabelecidos.

Cabe uma avaliação inicial e obtenção do entendimento da situação atual relativa aos dados existentes à luz das especificações de qualidade. A partir daqui é possível definir métricas para avaliar a qualidade dos dados. Este processo consiste inicialmente de uma avaliação geral dos dados atuais. O processo deve obter o grau de completude e o nível de acurácia do conjunto de dados avaliado.

A avaliação da qualidade da informação pode ser dividida em três níveis: atributo/campo (Tabela 5), tupla/registro (Tabela 6) e tabela/arquivo (Tabela 7).

Nível de atributo:

| Evidência: | Exemplo: |
|---|--|
| Valor ilegal, ou seja, fora do domínio de valores válidos (normalmente ocorrem em atributos numéricos e de enumeração). | Idade do cliente = 185 ; data entrada plano = “01/02/1969” quando a empresa foi criada em: “01/07/1970” |
| Nenhuma padronização de dados | R. Ouvidor, Rua do Ouvidor, Rua Ouvidor, Ouvidor R. |
| Valores inválidos | Atributo deve conter ‘A’ ou ‘B’ porem existem registros com valores = ‘C’ |
| Regras diferentes conforme cultura | Data = Jan 1, 2005 ou 1-1-2005 ou 1 Jan 05 |
| Formatos variados | (21)9674-2153 ou [21]96742153 ou 9196742153 |
| Padronização de nomes | jon j jones transformado em: Jon J Jones |
| Verificação | Código de endereçamento postal não corresponde à cidade/estado |
| Identificação de ausência de valor pelo preenchimento inadequado de atributos obrigatórios. Ou seja, valores obrigatórios são preenchidos aleatoriamente. | Telefone = “9999-9999” ; Data nascimento = ”01.01.1901” ou “11.11.1911” ou “1.1.1900” ou ainda “31.12.9999” |
| Erro ortográfico, associado a atributos textuais, decorrente de grafia errada e sem | cidade = “Rio de Jneiro” |

| | |
|--|--|
| nenhum critério de validação. | |
| Informação além do contexto da definição do atributo. | endereço residencial = “Rua do Carmo, nº 23, Centro - RJ / 20040-030” |
| Valor de significado indefinido, normalmente esta anomalia é decorrente da utilização de abreviaturas. | empresa = “MS” |
| Utilização de sinônimos e/ou expressões sintaticamente diferentes, mas iguais ao nível semântico. | profissão = ”professor” e profissão = ”docente” |
| Inexistência de uma representação padrão, ou seja, o valor do atributo surge sob formatos variados. | data = “04/05/2003” e data = “2003/05/04” |
| Definição de datas que assumem um dia como default, pois de fato só possuem informação de mês/ano. | Data da divulgação da cota previdenciária = “01/04/2008” Este valor sempre ocorre no primeiro dia do mês. |
| Descrição do significado do atributo incompatível com o seu conteúdo (ou sobreposição) | estado civil no registro de cliente deve constar: solteiro, casado, separado, divorciado e não se misturar com a classificação do cliente: prospect, cliente, bom cliente, ex-cliente. |
| Dado da coluna está sendo utilizado com várias finalidades. Nesses casos ainda pode-se encontrar: - Um espaço a frente do nome - Dois espaços entre o primeiro nome e o último nome e/ou entre os nomes. - Falta de ponto após a abreviação do nome. - Abreviação nos nomes do meio (às vezes abreviado, às vezes não). - Todos os caracteres em maiúsculo. - “&” ao invés de “e” para indicar vários | nome do cliente (definido como character (30)) e nele encontra-se caracteres especiais, tais como: % , \$, etc. |

| | |
|---|--|
| itens - “/” ao invés de “-“ | |
| Nomes não informados em campos obrigatórios. | Nome cliente=”Não Sei” ou “Não Definido”, “Não Disponível”, etc |
| Nomes que não existem na forma ortográfica utilizada e devem conter somente caracteres alfabéticos. | NOME=”J#se S8lva” Cidade=”C@lifornia” |
| Atributo que é válido pelo critério de validação de dígito verificador, porém indica que o mesmo não foi informado. | CPF=777777777-77 |
| Campos sobrecarregados: contêm informações sobre múltiplos e diferentes fatos armazenados. Isto acontece quando, ao invés de criar um novo campo para uma nova informação, se reaproveita um campo existente (campos sobrecarregados raramente são documentados). | NOME = “José Silva MORTO” se a pessoa já faleceu. |

Tabela 5 – Avaliação da qualidade da informação no nível de atributo

Nível de tupla/registros:

| Evidência: | Exemplo: |
|--|---|
| Um campo que foi definido como único em uma tabela, aparece identificado vários itens. Por exemplo a identificação de um produto, aparece identificando vários produtos. (normalmente por pertencerem a categorias diferentes) | Produto 1 = cama Produto 1 = geladeira |
| Violação da dependência entre atributos, gerando inconsistência entre os valores dos atributos. | idade = 30 e a data nascimento = “15/03/1979” |
| Violação da dependência entre atributos, gerando inconsistência entre os valores dos atributos, pela ausência de informação em um deles (seja com valor nulo ou valor inexistente). | cidade = “Rio de Janeiro” e código postal = “7777777” |
| Troca de valores entre atributos, ou seja, registro de | nome = “Rua A Quadra 2” e |

| | | | |
|--|--|-----------------|-------|
| valor no atributo errado. | endereço residencial = “José da Silva.” | | |
| Tupla vazia ou semi-vazia: a maioria dos atributos da tupla não se encontra preenchida. | | | |
| Transposição de palavras (ora é nome+sobrenome, ora é sobrenome+nome) | nome = “J.Santos” e nome = “Miller J.” | | |
| Translação de moeda (uso inadequado de um valor sem as casas decimais, que mal compreendido, leva à migração de valores maiores) | Mascara de edição | | |
| Referência existente, porém errada. Existe um valor no atributo que não viola a integridade referencial, mas que não é o correto (problema de atualização) | nome do funcionário = “Jose Alves” código departamento = 8 (porém este funcionário não trabalha neste departamento) | | |
| Definição de sexo (“F” ou “M”) para consumidores que conforme a cultura tanto podem ser homem como mulher. | Joan, Jan, Nicola e José pode ser homem ou mulher conforme a cultura e por isso a identificação do sexo se torna essencial e não passível de uma correção caso seja uma empresa multinacional. | | |
| Intervalos de data representados de forma incorreta | Emp | Data | Bonus |
| | 78175 | 15.3.2001 | 3500 |
| | 78175 | 15.3.2000 | 3250 |
| | 78175 | 12.3.1999 | 3140 |
| | 78175 | 2.1.1997 | 3000 |

Tabela 6 – Avaliação da qualidade da informação no nível de registros

Nível de tabela/arquivos:

Neste grupo incluem-se os problemas de qualidade dos dados que advêm de vários conjuntos de dados (tabelas, arquivos ou bases de dados), quando analisados como um todo.

| Evidência: | Exemplo: |
|--|---|
| Valor não único: duas entidades diferentes possuem valores iguais num atributo de valor único. | tupla de X: nº identidade=12121 e tupla de Y: nº identidade=12121 |
| Redundância sobre a mesma entidade, ou seja, a mesma entidade encontra-se representada sob formas iguais ou diferentes em várias tuplas. | data nascimento de X = “08/09/1973” e data nascimento de X = “25/01/1973” |
| Restrições de cardinalidade para um valor a ser referenciado. | Um novo produto que foi cadastrado, porém ainda não possui nenhuma venda para o mesmo. Espera-se encontrar no máximo três endereços para cada cliente. |
| Regras de medidas: existe um domínio razoável para um determinado valor numérico? | Valores de salário estão entre > 0 e $< R\$50.000,00$ (Se for < 0 ou $> 50.000,00$ caracteriza-se em erro) Ofertas de produto estão entre > 1 e $< 1.000,00$ (Se for < 1 ou $> 1.000,00$ caracteriza-se em erro) |
| Coluna que só contém dados se existe valor em outra coluna ou não. | Na tabela cliente, alguns clientes são pessoas jurídicas e outras pessoas físicas. Caso seja pessoa jurídica, possuem CNPJ, tamanho da organização, etc. |
| Acompanhamento do crescimento da tabela/arquivo | A tabela X cresce em média 1000 linhas por mês. Entre março de 2001 e dezembro de 2001 a tabela Y |

| | |
|--|--|
| | <p>cresceu de 10000 para 12000 por mês. Em janeiro de 2002 somente 1/3 das linhas do mês anterior são esperadas.</p> |
|--|--|

Tabela 7 – Avaliação da qualidade da informação no nível de arquivo

A avaliação dos dados permitirá obter uma perspectiva da situação atual dos dados na organização e irá contribuir para descobrir a qualidade, características e problemas em potencial de informação antes mesmo do início de alguns projetos. Reduz drasticamente o tempo e os recursos necessários para encontrar dados problemáticos; permite aos analistas de negócio e aos administradores de dados terem maior controle na manutenção e administração dos dados corporativos. Permite catalogar e analisar metadados e descobrir relações entre metadados.

4.5.2 Sincronização de Dados

Nessa fase a Governança de Dados deve tratar de procedimentos que devem ser estudados minuciosamente para manter as soluções de integração e consistência dos dados. Lembrando que Sincronização de dados é o processo que garante a integração e consistência de dados em tempo real para vários aplicativos ou bancos de dados na organização. A política de sincronização de dados, nos diversos patamares da arquitetura de informações de uma organização, está diretamente relacionada à governança dos dados, garantindo assim qualidade na disponibilização única da informação. O importante é possibilitar a integração de diversas informações de negócios tanto da empresa como de seus parceiros. Soluções para integração orientada à informação podem ser encontradas em: LINTHICUM (2003) e ZIEGLER (2007).

4.5.3 Segurança dos dados

Nessa fase a Governança de Dados deve tratar dos procedimentos que devem ser estudados minuciosamente para proteção e segurança interna dos dados. A análise inicia-se pela definição do Firewall na rede, bloqueando portas não privadas, deixando aberto publicamente apenas os serviços comuns (HTTP, HTTPS, smtp, pop, imap). O Banco de dados deve ter acesso controlado, sendo liberado apenas para IPs previamente cadastrados.

A governança também deve estabelecer um esquema de classificação de dados aplicável a toda a organização com base na importância e na confidencialidade dos dados corporativos (por exemplo: público, confidencial, altamente secreto). Esse esquema inclui detalhes sobre os proprietários dos dados, definição de níveis apropriados de segurança, controle de proteção, breve descrição dos requisitos de retenção e destruição dos dados, importância e confidencialidade. É utilizado como base para aplicação de controles, tais como controles de acesso, arquivamento ou criptografia.

Também é pertinente ter registros de acessos (Log) para todas as alterações feitas para a maioria dos serviços que envolvem segurança de dados (FTP, Painel de Controle, manutenção de e-mails), ficando registrado o Internet Protocol (IP) utilizado pelo autor das mudanças, bem como data e horário da alteração e outras informações relativas ao acesso/alteração feita pelo usuário.

No que diz respeito ao Backup, todo o conteúdo do banco de dados deve ser copiado (backups) diariamente e as cópias armazenadas em pontos geográficos diferentes, garantindo a máxima segurança no caso de catástrofes, erros de servidor ou simples perda por "engano" do usuário. Como medida máxima de segurança, recomenda-se manter sempre uma cópia atualizada de seus dados em sua organização, possibilitando uma recuperação rápida e imediata para ocorrências recentes.

No que diz respeito à segurança física, os servidores devem encontrar-se alojados dentro dos melhores e mais seguros centros de dados, com rigoroso controle de temperatura, controle de umidade e fluxo de ar. Estas salas devem estar equipadas com detectores e sistemas de combate ao fogo e protegidos contra furacões, inundações e outros fenômenos da natureza. A vigilância deve ser 24hs, através do monitoramento e gravação das imagens das atividades na mesma. A movimentação de funcionários e/ou visitantes deve ser controlada por equipe de segurança, que mantém registro de horários das visitas e identificação dos visitantes. O tratamento da rede elétrica, através da utilização de *NoBreak* e geradores, que devem garantir a comutação automática e a continuidade do ambiente, em caso de corte de fornecimento pela concessionária até que sejam executados os procedimentos normais de desligamento dos mesmos, no caso de não haver previsão de restabelecimento.

4.5.4 Padronização de Dados

A padronização de dados inclui padrões de nomes de elementos de dados, especificação de requisitos, modelos de dados, projeto do banco de dados e arquitetura e procedimentos para cada função da administração de dados. Normalmente esses padrões são definidos pelo profissional de administração de dados e devem ser revisados, aprovados e adotados pelo comitê de Governança de Dados. Esses padrões e orientações devem ser amplamente divulgados, monitorados, obrigatórios e revisados periodicamente.

Padrões e diretrizes procedimentais de dados devem incluir:

- Padrões de modelagem de dados e arquitetura, incluindo convenções de nomes de elementos de dados, padrões de definição, padrões de domínio e padrões de abreviação.
- Padrões de negócio e metadados técnicos para capturar, manter e integrar dados.
- Procedimentos e diretrizes da administração do modelo de dados.
- Integração de metadados e procedimentos de utilização.
- Procedimentos para recuperação de banco de dados e continuidade de negócio, desempenho do banco de dados, retenção dos dados e aquisição de dados externos.
- Diretrizes e procedimentos para segurança de dados.
- Diretrizes e procedimentos para unificação e limpeza de dados.

Nas empresas, há certas normas que podem ser aplicadas a cada “pedaço” de dados. Por exemplo, o nome e o endereço deve sempre obedecer às normas dos serviços postais. Endereços de email em conformidade com uma determinada forma (nome de usuário, símbolo "@", domínio de internet). No entanto, podem haver dados que a equipe de governança precisa definir: um novo padrão, ou seja, regra de numeração, a descrição de item, dados da cadeia de fornecimento e outros dados não tratados anteriormente. Para isto, a definição deve surgir em conjunto com a equipe de negócios. Como parte deste processo, explorar os dados atuais ajudará a decidir quais os dados que existem e encontram-se corretos, em especial em campos obrigatórios, e estabelecer normas de sistema que podem ser automatizadas e controladas para cumprimento de exigências legais.

4.6 Oportunidades e soluções

Nesta quinta fase da proposta deste trabalho, cabe enumerar as várias possibilidades de implementação, identificando os principais projetos de implementação que poderiam ser assumidos e avaliando a oportunidade para o negócio associada a cada uma delas. O padrão TOGAF recomenda que o primeiro passo seja dar enfoque aos projetos que produzirão benefícios em curto prazo e, assim, criar uma força propulsora para dar andamento a projetos de prazo mais longo. Essa é uma boa recomendação para qualquer metodologia arquitetural (THE OPEN GROUP, 2009). Assim sendo, deve-se procurar as ações que possam ser realizadas da forma mais econômica possível e que, ao mesmo tempo, produzam o mais alto valor observado. Deve-se também buscar o provimento de soluções tecnológicas seguras, eficazes e consistentes em toda a organização, estabelecendo um fórum de tecnologia para prover diretrizes tecnológicas, aconselhamento sobre a infra-estrutura de produtos, orientação na seleção da tecnologia e avaliação da conformidade com estes padrões e diretrizes. Este fórum direciona os padrões e práticas tecnológicas com base na relevância para o negócio, nos riscos e na conformidade com requisitos externos.

4.7 Plano de implementação

Neste momento deve-se definir a metodologia e coordenação das atividades relacionadas à implementação das oportunidades e soluções apontadas na fase anterior, bem como a opção pelo fornecimento de ferramenta para execução, se for pertinente. Caso haja necessidade de conversão de dados, cabe à equipe planejar, coordenar e executar as atividades de extração e carga da base de dados, incluindo:

- mapeamento das informações dos sistemas legados para o novo modelo de dados;
- definição da estratégia de migração de dados;
- execução e testes das extrações;
- cargas de dados convertidos no novo ambiente;

Após a definição do modelo de dados, ele deve ser devidamente documentado com o domínio, tamanho, tipo e formato de cada atributo, juntamente com as regras de integridade.

Também é necessário realizar um levantamento dos impactos relativos à migração da base atual de dados, com a finalidade de adequar ao cronograma o volume de informações a serem migradas. Caso sejam necessários recadastramentos ou levantamentos em campo para atualizar ou obter informações inexistentes nos sistemas legados, o comitê de Governança de Dados da empresa será responsável pela identificação desses recursos e orientação para execução dessas atividades.

Aqui cabe a entrega de um Plano de Migração de Dados, Solução para Extração dos Dados e Análise de Impacto da Migração de Dados.

4.7.1 Cultura e comportamento em relação à informação

O comportamento informacional pode ser definido como a maneira como os indivíduos lidam com a informação, compreendendo a busca, o uso, a alteração, a troca, o acúmulo e o ato de ignorá-la. Já a cultura informacional é o comportamento informacional de grupos ou de organizações, abrangendo valores, crenças, se orientada por fatos ou por rumores, com enfoque interno ou externo, controlado ou disseminado, entre outros. Existem várias razões para a administração dessa cultura. As mais importantes são que o valor de uma empresa é afetado pelo conhecimento que ela possui e a redução de custos, pelo compartilhamento de informações compradas externamente.

O compartilhamento de informações deve estar apoiado na cultura organizacional e para isso a empresa deve reconhecer e incentivar a disponibilização de informações. Entretanto, deve-se tomar certo cuidado com a sobrecarga de informações. A informação deve ser disponibilizada para a pessoa certa, que deve, por sua vez, reconhecê-la e utilizá-la.

4.7.2 Treinamento

Existe a necessidade de se organizar treinamentos em vários níveis. Para habilidades gerais e para algumas habilidades específicas de domínio, é possível encontrar programas disponíveis externamente. Para habilidades mais específicas da organização, é necessário desenvolver e planejar apresentações, workshops e, em alguns casos, programas de treinamento mais específicos.

Entende-se por habilidades gerais, por exemplo, promover a formação que não somente oriente a processos e a seguir procedimentos, mas também ao atendimento e satisfação de seu cliente. Outro exemplo de habilidade geral é proporcionar a todos a compreensão das regras de negócios para tomar as decisões corretas. Outras habilidades

específicas de domínio são, por exemplo, a formação com noções básicas sobre o negócio, incluindo produtos e serviços, e onde estão envolvidos os atores de negócios (clientes, parceiros, fornecedores); o entendimento de que resultados são produzidos e quais serviços são entregues e como isto se encontra relacionado às responsabilidades de trabalho dentro dos processos.

Já as habilidades específicas de processos de negócios podem ser:

- Um bom conhecimento do processo ou processos em que a pessoa está inserida.
- Um bom conhecimento das responsabilidades definidas para os trabalhadores de negócios, papel e a capacidade de executar essas atividades.
- Uma compreensão das responsabilidades e das atividades dos colegas.
- Conhecimento do uso das ferramentas de negócios.

Obtenção das habilidades corretas dentro da organização pode ser uma combinação de treinamento de funcionários com contratação de pessoas.

Pode-se definir um sistema de premiação para estimular funcionários a trabalhar na direção da idéia e da estratégia do negócio. Com as metas de processos de negócios individuais, que como um ponto de partida, devem ser baseadas em estratégias e idéias de negócios, definir as premiações relacionadas a:

1. Desempenho geral do negócio.
2. Desempenho geral do processo de negócio.
3. Resultados da execução individual (instância) de um processo de negócios.
4. Contribuições de cada indivíduo.

4.8 Implementação da governança

Na sétima fase deste framework prevê-se a divulgação do andamento do programa de Governança de Dados através de um processo de distribuição formal, sendo disponibilizado em um portal da empresa ou na intranet ou *newsletter* de forma que todos tenham acesso às atualizações, métricas e realizações alcançadas. Uma estratégia de comunicação deve ser bem delineada, determinando que todos na organização devem ser comunicados, que informações são importantes, que frequência deve ser feita a atualização dessa informação e qual o método mais eficiente de disseminação da informação na empresa.

Definir e comunicar para todos os usuários informacionais os seus respectivos papéis e responsabilidades, especificando a autoridade, responsabilidade e responsabilização, com o objetivo de atender às necessidades da organização. É importante um programa contínuo de monitoramento para assegurar que a qualidade dos dados está se mantendo em níveis satisfatórios e atende às necessidades organizacionais.

4.9 Administração da governança

Conceitualmente, mudança é uma ação (adição, modificação ou remoção) que resulta em uma nova situação de um ou mais itens de configuração. Para que uma mudança tenha êxito e seja permanente, também é necessário compreender, e possivelmente mudar, a cultura da organização. Caso não haja compreensão sobre a cultura da organização, qualquer esforço de governança buscando a excelência da qualidade dos dados falhará. A cultura não é algo palpável ou que possa ser descrito em uma fórmula simples e deve ser tratada de forma a não trazer elementos que interfiram na organização de forma inesperada. A gestão da mudança é um processo eficaz de alocação de recursos, de forma a transformar a organização, com o objetivo de melhorar a sua eficácia. Existem diversas condicionantes que influenciam o processo de mudança de uma organização, como o nível de recursos que a empresa dispõe, o setor onde se insere, a maior ou menor necessidade de mudança, qualidade da gestão e a política, econômica, social, tecnológica, ambiental e legal. Mesmo a equipe gerencial mais bem-intencionada e respeitada, se não alcançar a adoção entusiástica das mudanças, verá seu plano se reverter nas rotinas velhas e familiares. Sem um sólido programa de mudanças gerenciais perfeitamente entrelaçado com a Governança de Dados, assim como uma forte liderança, a governança estará condenada desde o seu início.

Segundo John Kotter (2005), existem 8 etapas para um processo de mudança bem sucedido:

1. Criar um sentido de urgência. – Ajude aos outros a verem a necessidade de mudança e a importância de agir de imediato.
2. Trabalhar em equipe. – Certifique-se de que há um grupo forte orientando a mudança, com capacidade de liderança, linhas de ação, credibilidade, capacidade comunicação, autoridade, e capacidade de análise.

3. Desenvolver a visão e a estratégia da mudança. – Esclareça como o futuro será diferente do passado, e como se pode fazer que o futuro seja uma realidade.
4. Comunicar para melhorar a compreensão. – Certifique-se que é possível compreender e aceitar a visão e a estratégia.
5. Motivar e capacitar àqueles que querem participar. – Remova todas as barreiras possíveis, para aqueles que querem ajudar a tornar uma realidade a visão e afastar os que estão contra o processo.
6. Obter pequenas vitórias. – Produza alguns sucessos com a maior brevidade possível, ainda que não sejam consolidados, mas permita credibilizar o processo da mudança. Comece por ganhar algumas batalhas, para depois vencer a guerra.
7. Não desistir. – Seja exigente e mais rápido após os primeiros sucessos. Seja implacável com a instituição da mudança passo a passo, obstáculo a obstáculo, até a visão se tornar uma realidade.
8. Criar uma nova cultura. – Assegure a aplicação das novas formas de comportamento, e certifique que se conseguiu, até se confundirem com a própria cultura da organização, na qual se expressa o modo de vida, o sistema de crenças e valores e a forma de interação e de relacionamento típicos de determinada organização.

4.10 Administração de requisitos

Conceitualmente, define-se requisito como uma condição ou uma capacidade que um sistema deverá apresentar. O gerenciamento de requisitos é uma abordagem sistemática para levantar, documentar, organizar e controlar os requisitos variáveis em um sistema. O desafio da administração de requisitos reside no fato de que os requisitos são dinâmicos e as alterações podem causar impacto na governança e/ou nas estruturas de dados. O correto gerenciamento dos requisitos é condição fundamental para o sucesso do projeto, bem como na gestão dos dados, na medida em que busca também firmar e atualizar acordos entre o cliente e a equipe do projeto sobre os requisitos variáveis do sistema. Os pontos críticos para o gerenciamento eficaz de requisitos são manter uma definição clara dos requisitos, juntamente com os atributos aplicáveis e a rastreabilidade para outros requisitos e outros artefatos do projeto.

A coleta de requisitos pode parecer uma tarefa bem precisa. Na realidade, porém, os projetos enfrentam dificuldades pelos seguintes motivos:

1. nem sempre os requisitos são óbvios;
2. podem vir de várias fontes;
3. nem sempre são expressos em palavras de modo fácil ou claro;
4. existem diversos tipos de requisitos em diferentes níveis de detalhe;
5. uma quantidade significativa de requisitos não controlados pode inviabilizar a gerência de escopo;
6. os requisitos estão relacionados uns com os outros, e também com o produto liberado do processo de engenharia do software;
7. os requisitos têm propriedades exclusivas;
8. existem várias partes interessadas, o que significa que os requisitos precisam ser gerenciados por grupos de pessoas de diferentes funções;
9. e os requisitos são mutáveis.

Desta forma, diversas habilidades são necessárias para o gerenciamento dessas dificuldades. As habilidades abaixo relacionadas são importantes para o gerenciamento de requisitos:

- Analisar o problema;
- Identificar as necessidades dos envolvidos;
- Definir as características do sistema;
- Definir as características dos dados necessários;
- Gerenciar o escopo do projeto;
- Refinar a definição do sistema; e,
- Gerenciar as mudanças nos requisitos.

Cabe ao levantamento de Requisitos de dados:

- Estabelecer e manter concordância com os clientes e outros envolvidos sobre o que os dados devem representar;
- Oferecer aos envolvidos com a manipulação e uso da informação uma compreensão melhor dos requisitos de dados;
- Definir as fronteiras dos dados (ou delimitar o escopo);
- Fornecer uma base para planejar o conteúdo técnico das iterações;
- Fornecer uma base para estimar o custo e o tempo de garantir a qualidade de dados;
- Definir uma interface de usuário para acesso aos dados, focalizando nas necessidades e metas de qualidade de dados dos usuários;

Para atingir essas metas, é importante, antes de tudo, compreender a definição e o escopo do problema a ser resolvido. Os produtos resultantes da Modelagem de Negócios são fontes de informação a serem consultadas durante o trabalho de definição dos requisitos de dados. Para tal deve-se utilizar a modelagem de dados que é um método usado para definir e analisar requisitos de dados necessários para apoiar o processo de negócio de uma organização. Os requisitos de dados são registrados em um modelo de dados conceitual com as definições dos dados e seus tipos. Implementar um modelo de dados conceitual pode requerer vários modelos de dados lógicos. Por fim, os modelos lógicos são transformados em modelos físicos, representados em alguma linguagem ou banco de dados. A modelagem de dados define não apenas seus elementos, mas sua estrutura e seus relacionamentos e pode ser realizada em vários tipos de projetos e em diversas fases do projeto. A definição destas características envolve membros da equipe de Governança de Dados e representantes do cliente. O levantamento dos requisitos de dados de um projeto de Governança de Dados concentra-se prioritariamente nas Fases 1 (Iniciativas Estratégicas) e 2 (Visão, escopo e objetivos) da proposta de Governança de Dados. No entanto, uma vez que a governança é realizada de forma iterativa, é natural que esta atividade se estenda ao longo das demais fases. Assim, torna-se possível tratar as mudanças nos requisitos e as eventuais alterações de escopo. Cabe ao Analista de Requisitos refinar os requisitos de qualidade de dados a serem desenvolvidos, e definir como a funcionalidade deverá ser implementada em termos de colaboração entre objetos.

5. O Estudo de Caso

A Empresa descrita a seguir foi utilizada como base do estudo de caso, considerando-se um escopo reduzido, possibilitando assim utilizar a metodologia proposta neste trabalho na sua totalidade, dentro do prazo possível para o desenvolvimento deste trabalho e permitindo sua avaliação. Pelo fato do estudo centrar-se em aspectos de Governança de Dados, foram envolvidos tanto profissionais da área de TI como da área de negócio, mais especificamente ligados à atividade fim da empresa.

5.1 Caracterização da Empresa

A fim de fornecer uma visão geral sobre o ambiente de realização deste estudo, a organização que foi objeto de estudo será brevemente caracterizada. Por motivo de confidencialidade, a empresa é tratada como Empresa X. Apesar disso, algumas características fundamentais da organização são apresentadas para possibilitar a compreensão do contexto em que ocorreu o processo de aplicação da metodologia de Governança de Dados e porque algumas decisões consideradas chaves foram tomadas. A Empresa X é uma empresa de médio porte, sediada no Rio de Janeiro, que atua na área de Previdência Complementar. Quando iniciou suas atividades, há quarenta anos, a Empresa X somente comercializava um plano de previdência, restrito aos funcionários de algumas empresas de um único conglomerado. A empresa cresceu, acompanhando a estabilidade da economia brasileira e conseqüentemente o crescimento do mercado de previdência complementar. Atualmente ela continua atuando neste ramo, oferecendo vários planos para várias empresas atuantes em diversos ramos de negócio.

Sua missão é *"Oferecer produtos e serviços, com foco em seguridade, adequados às expectativas de seus participantes e patrocinadores, geridos com eficácia, transparência e responsabilidade social"*. Sua base de clientes hoje gira em

torno de 140 mil pessoas entre ativos e assistidos e nela trabalham cerca de 400 funcionários.

5.2 Caso da Empresa X

Em 2007 foi executado um projeto de avaliação da qualidade dos dados das informações de clientes e promovendo o saneamento das mesmas, patrocinado pela Gerência de Tecnologia. Na ocasião, foram detectados vários problemas, dentre os quais, o mais grave a duplicidade de registros. Além do saneamento, alguns mecanismos criados e aplicativos foram alterados de forma a minimizar esse problema.

Cabe ressaltar que nem todos os problemas foram devidamente sanados, por ser necessária uma avaliação da área de negócio, que não tinha o comprometimento devido com o trabalho em questão. A principal alegação é que o mesmo havia sido contratado pela TI e como eles possuem outras atividades e metas, não podiam priorizar esse assunto. Na ocasião foi sugerido que a base de dados de clientes tivesse uma única área gestora e responsável pela atualização dessas informações e que a mesma fosse validada junto à Secretaria de Receita Federal, certificando-se assim a veracidade do CPF informado. Foram implementadas algumas regras no banco de dados a fim de garantir a unicidade do CPF, porém não se conseguiu torná-lo obrigatório, uma vez que a área de negócio alega que um menor de idade não é obrigado a possuir CPF (a não ser que ele se torne um recebedor de um benefício). O nome abreviado, definido com 20 posições e que tem por objetivo agilizar possíveis consultas, onde anteriormente era cadastrado aleatoriamente pelo usuário, foi padronizado segundo regra definida pela área financeira e automatizada por uma rotina de formatação. Apesar de todos os esforços, alguns registros não foram acertados e nem outras ações que se faziam necessárias foram tomadas.

5.3 Utilizando o framework proposto na Empresa X

Para o trabalho em questão, a mesma base de dados de clientes foi utilizada. Foi assumido que a partir do Planejamento Estratégico, surgiria a diretriz de definir estratégias de Governança de Dados objetivando a obtenção de maior qualidade nos produtos e serviços oferecidos aos participantes, mantendo assim o alinhamento com a

Missão da organização. Constando também das *Iniciativas Estratégicas*, foi assumida a aprovação da formação de um comitê de política e gestão de dados e a orientação para definição de políticas e procedimentos que serão adotados em toda a organização. Esse comitê foi inicialmente formado por representantes das Gerências de Benefícios e de Tecnologia da Informação, bem como da Assessoria de Planejamento. A justificativa para o mesmo se baseia no reconhecimento da alta direção de que as medidas de desempenho e administração são fundamentais para o sucesso da organização, e que são diretamente dependentes de dados precisos e confiáveis. Ainda reconhecem que a boa qualidade da informação é essencial para administrar e planejar atividades bem como apoiar o processo decisório.

A partir de então, iniciou-se a primeira iteração proposta neste trabalho (Figura 23), pela definição da primeira etapa, *Visão e Objetivos*, definindo a intenção de melhorar as informações referentes aos dados pessoais dos clientes participantes de Plano de Previdência Complementar, objetivando o atendimento da instrução SPC nº 26, que orienta às EFPC (Entidades Fechadas de Previdência Complementar) a manter permanentemente atualizadas as informações cadastrais de seus clientes, a saber:

- Nome completo, sexo, data de nascimento, naturalidade, nacionalidade, estado civil, filiação e nome do cônjuge;
- Seu enquadramento na condição de pessoa politicamente exposta, se for o caso. (A pessoa encontra-se nesta condição caso desempenhe ou tenha desempenhado, nos últimos cinco anos, cargos, empregos ou funções públicas relevantes, no Brasil ou em outros países e territórios estrangeiros e em suas dependências, bem como seus representantes, familiares e outras pessoas de seu relacionamento próximo).
- Natureza e número do documento de identificação, nome do órgão e data de expedição.
- Número de inscrição no cadastro de pessoas físicas (CPF)
- Ocupação profissional principal

Também ficou definido que os critérios descritos a seguir seriam utilizados para mensurar e divulgar, objetivando a obtenção do máximo de qualidade dos dados.

- Perfeição: Todo dado relevante ou requerido está prontamente disponível para utilização quando necessário.

- **Acurácia:** Dados estão corretos e precisamente refletem o objeto ou a transação que está descrita.
- **Confiabilidade:** Dados são ou se mantêm consistentes através das várias transações de sistemas.

Continuando a orientação do framework, a equipe partiu para a análise dos *Processos de Negócio*, tanto no que diz respeito à forma como essas informações chegam à organização e como elas são mantidas. Com isso foi necessário rever o protocolo de relacionamento firmado entre a Empresa X e as empresas patrocinadoras, bem como efetuar a revisão do manual de interface entre as mesmas, passando a exigir-se que os dados relacionados na instrução SPC nº 26 fossem enviados.

Outra definição oriunda das análises feitas, diz respeito à padronização dos Pedidos de Inscrição (PI) para todos os Planos de Previdência, tornando obrigatório o preenchimento das informações exigidas na instrução SPC nº 26. Com relação ao processo de deferimento do PI, uma vez que o mesmo deveria ser entregue assinado, ficou determinado que o cliente passasse a enviar cópia dos documentos por ele informados, ou seja, carteira de identidade, CPF, e comprovante de ocupação, que serão digitalizados na Empresa X, juntamente com o próprio PI, fazendo parte do dossiê do participante.

Na etapa seguinte do framework, que se refere a *Sistemas de Informação*, coube uma revisão do modelo de dados, definição das métricas de qualidade de dados e das regras de negócio. No que diz respeito à modelagem de dados, a mesma foi revista não necessitando nenhum ajuste, uma vez que todos os atributos já se encontrariam devidamente definidos. Porém no que diz respeito ao tratamento desses atributos pelo cadastramento de Pessoa Física existente hoje no sistema, as definições de obrigatoriedade e as regras de negócio precisariam ser alteradas, tornando obrigatórios os preenchimentos das informações em questão. A área de negócio também apontou como necessidade urgente que essas rotinas de validação e obrigatoriedade fossem implementadas de forma a serem reutilizadas pelo aplicativo de preenchimento do PI Online, mantendo-se assim o mesmo critério.

Na etapa de *Tecnologia e Padronização*, algumas iniciativas foram tomadas, como por exemplo, unificação do formulário de Pedido de Inscrição, onde passariam a constar as informações exigidas pela Instrução.

Também se executou um trabalho de padronização, limpeza e levantamento dos possíveis Órgãos de Expedição da Identidade. Foi adotada uma relação de nomenclaturas de Órgãos e Instituições que emitem documento de identidade, não ficando mais a cargo do usuário sua digitação. No que se refere ao CPF, esta informação passaria a ser validada pela cópia do documento enviado. Caso haja alguma dúvida, o setor que faz o deferimento do PI deverá validar no site da Secretaria de Receita Federal / Cadastramento de Pessoa Física.

Com relação às informações de Pessoas Politicamente Expostas, estas serão mantidas atualizadas a partir do cadastro do Serasa, órgão responsável por mantê-lo atualizado. Já a ocupação principal passa a ser validada pelo Cadastro Brasileiro de Ocupações, mantido pelo Ministério do Trabalho e Emprego - MTE. Este cadastro foi instituído por portaria ministerial nº. 397, de 9 de outubro de 2002 e tem por finalidade a identificação das ocupações no mercado de trabalho, para fins classificatórios junto aos registros administrativos e domiciliares.

Como sugere o framework, nas *Oportunidades e Soluções* seriam resgatadas algumas informações inexistentes na base de dados, a partir da base de dados do sistema legado, corrigindo assim algumas ausências percebidas. Além disso, foram também levantados os documentos de identidade que se encontram digitalizados para pessoas que não possuem essa informação na base. O comitê sugeriu a utilização de estagiários do setor de Benefícios para consultar esses documentos, pelo software que administra imagens, resgatando e completando essas informações pelo aplicativo de cadastramento de Pessoa Física da Empresa X.

No *Plano de Implementação* foi definido como prioridade os ajustes necessários nos aplicativos, de forma a passar a tratar a informação o mais breve possível. E também foi estabelecida uma equipe que estaria promovendo as correções pertinentes na base dados imediatamente. Também foi estabelecida nesta etapa proporcionar treinamento a todos os colaboradores, promovendo assim uma mudança cultural, através da unificação de conhecimento sobre o tratamento dessas informações.

Coube à *Implementação da Governança* acompanhar todas as implementações efetuadas, garantindo que as alterações serão gerenciadas de forma coesa e projetada. Cumprindo a determinação da administração, a divulgação dos resultados obtidos para toda a organização será feita através da Gerência de Comunicação Institucional utilizando o canal de voz existente atualmente na empresa, bem como a página na Intranet.

Como atividade da fase de *Administração da Governança*, o comitê definiu que esse canal de voz também seria utilizado para dar dicas de comportamento em relação à informação, através de chamadas eventuais e aleatórias, em horários diferentes, procurando assim atingir o maior número possível de profissionais desta empresa.

5.4 Avaliação do Método

Nesta seção encontram-se descritos os resultados da avaliação do estudo de caso realizado na Empresa X. Após apresentação da proposta de Governança de Dados e sua utilização, conforme descrito anteriormente, o grupo de 9 (nove) pessoas envolvidas nesta primeira iniciativa de governança na organização foi convidado a fazer uma avaliação, respondendo ao questionário que se encontra no Anexo II.

5.4.1 Perfil dos Respondentes

Os participantes desta avaliação estão, em geral, há bastante tempo na organização: 55% estão há mais de 20 anos, 22% encontram-se na mesma entre 10 e 20 anos e 23% estão entre 1 e 5 anos na organização.

Quanto à formação acadêmica, os respondentes apresentam um perfil bem equilibrado em termos das duas visões (TI e negócios) cujo alinhamento foi discutido nesta dissertação. 45% possuem formação em Tecnologia da Informação, 45% possuem formação em Administração de Empresas e 10% em Economia. Apesar disto, 70% dos respondentes ocupam o cargo de Analistas de Sistemas e 30% de Gerência de Negócio. Em relação à área de atuação, 50% se classificam atuantes na área de sistemas; 30% se classificam atuantes na área de Dados e 40% se classificam atuantes na área de Negócio (Figura 25).

Cabe ressaltar que os gerentes se classificaram como atuantes na área de negócio e dados, o que denota uma conscientização de sua responsabilidade nessas duas áreas. Já os 50% que se enquadraram na área de sistemas são Profissionais de TI, todos da área de Análise de Sistemas. Como analistas de sistemas, estes profissionais deveriam entender que sua atuação também é fundamental na área de negócio e dados. Isto denota uma falta de maturidade dos profissionais de TI em relação a sua responsabilidade nessas áreas de atuação, uma vez que Sistemas de Informação contempla tecnologia,

procedimentos e pessoas. Para desempenhar seu papel a contento, o Analista de Sistemas necessita ter o conhecimento do negócio bem como perceber a necessidade de informação envolvida no processo de negócio.

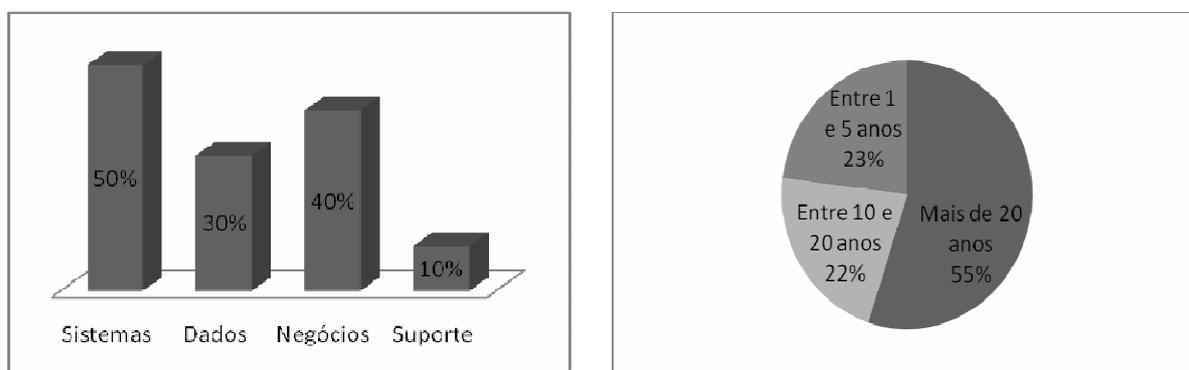


Figura 25 - Área de Atuação dos respondentes e tempo na organização

5.4.2 Resultados obtidos

Todos os entrevistados foram unânimes em afirmar que a participação, engajamento e direcionamento da alta administração no apoio às tarefas relacionadas à Governança de Dados é vital para o seu êxito. Além disso, 20% dos respondentes assinalaram que além de vital, também é fundamental essa participação.

Um dos respondentes da área de negócios acrescentou que considera fundamental que a alta administração patrocine um processo de mudança cultural, a fim de informar para toda a empresa a relevância da qualidade da informação. Outro da mesma área acredita que a confecção de uma norma para fins de nortear e garantir a execução do processo seja fundamental, garantindo assim que, independente de quem esteja na direção da organização, a execução da Governança de Dados estaria garantida por esta norma.

Cabe ressaltar que é uma característica da Empresa X haver grandes mudanças nos cargos da alta administração, bem como no Conselho Deliberativo e Fiscal, a cada novo mandato da presidência da República e conseqüentemente da direção da empresa patrocinadora e fundadora desta instituição.

Já um profissional de TI acrescentou que “acha muito difícil qualquer projeto, programa, ou atividade que exija participação corporativa conseguir ter êxito sem o apoio da alta administração”; outro acrescentou que “esse patrocínio é fundamental para priorizar as atividades na empresa”. E ainda segundo a visão de outro profissional de TI,

“havendo o comprometimento da alta administração fica mais fácil consegui-lo nas outras esferas”.

Do conjunto de pessoas que responderam ao questionário, 90% pensam que a Governança de Dados proposta que utiliza uma visão através de um framework corporativo, pode ter uma maior adesão e receptividade, uma vez que a governança está explicitada de forma clara e concisa (Figura 26). Já 80% das pessoas desse mesmo grupo acreditam que o framework de Governança de Dados proporcionará uma maior conscientização dos CEOs de como a questão deve ser tratada na organização, e 65% também acreditam que o framework promoverá uma discussão interna motivadora sobre a Governança de Dados. Somente um respondente da área de negócio da Empresa X fez questão de destacar que o framework corporativo apresentado aos CEOs da empresa poderá provocar conflitos de interesse e poder entre os executivos da organização. Não foi possível identificar o motivo, uma vez que o mesmo não detalhou, apesar de existir essa possibilidade no questionário. Seria necessário explorar mais esta questão através de uma entrevista específica com este respondente. As demais opções existentes para esta questão não foram destacadas por nenhum questionado.

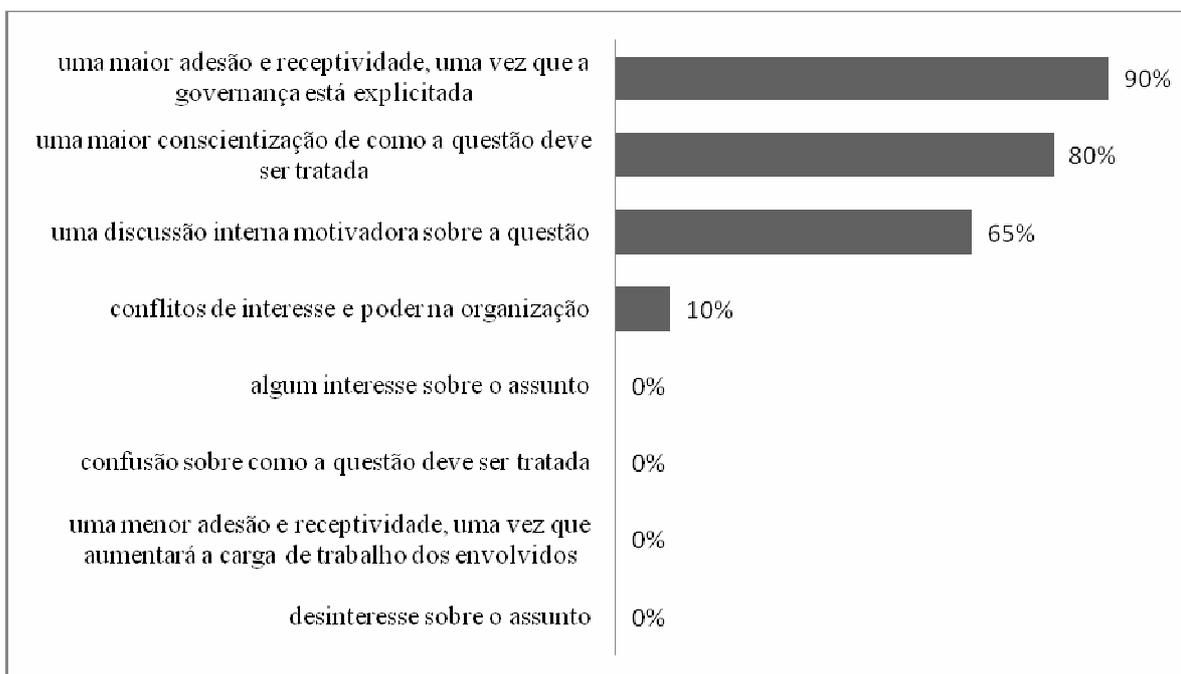


Figura 26 - Influência na organização ao utilizar um Framework Corporativo de Governança de Dados

Quanto à avaliação da complexidade de entendimento das etapas da proposta de governança apresentada baseada no TOGAF-ADM, 60% dos profissionais que

responderam a pesquisa não encontraram dificuldade para entender as etapas da proposta de Governança de Dados apresentada. Ainda 20% desse grupo de profissionais perceberam dificuldade de entendimento tanto na etapa de *implementação* quanto na de *administração da governança*, por acreditarem que essas etapas estão diretamente ligadas à questão do patrocínio e apoio da alta administração na implementação da governança.

Dos respondentes deste questionário, 70% acham que a etapa mais importante da proposta de Governança de Dados baseada no TOGAF-ADM é a primeira etapa, ou seja, *Iniciativas Estratégicas*. Em torno de 60% do grupo acha que em segundo lugar em nível de importância fica a etapa de *Administração da Governança*. Em terceiro lugar em nível de importância, 45% do grupo destacou a fase de *Implementação da Governança*. Já as etapas: *Tecnologia e Padronização* e *Administração de Requisitos* ficaram em quarto lugar em importância para cerca de 35% dos respondentes. Somente 25% dos respondentes colocariam em quinto lugar em nível de importância, as etapas: *Visão e Objetivos* e *Oportunidades e Soluções* (Figura 27).

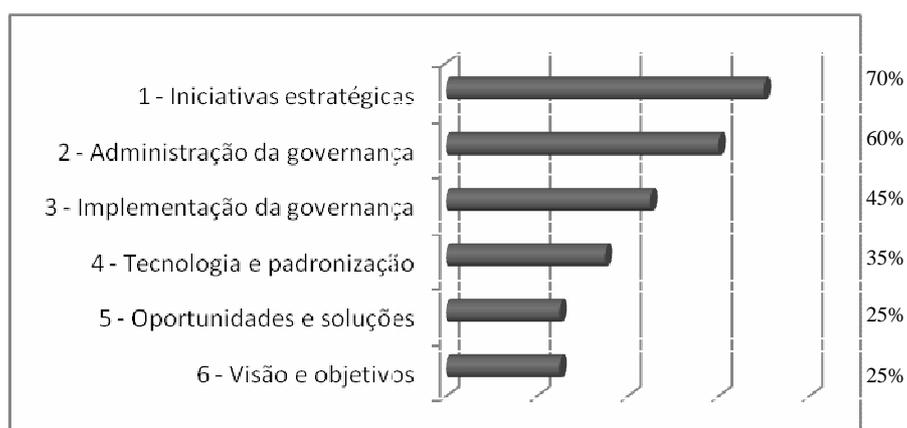


Figura 27 - Ordem de importância das Etapas de Governança de Dados

Percebe-se que o grupo destacou as *Iniciativas estratégicas*, *Administração* e *Implementação da governança* por acreditar ser primordial o apoio da alta administração e a definição de diretrizes e estratégias, essenciais à iniciativa e a continuidade da Governança de Dados. Um respondente enfatizou que, no seu entendimento, tudo começa com a iniciativa do ponto de vista estratégico, onde as decisões são tomadas, porém isto deve estar alinhado com o *plano de implementação* para fins de controle e obtenção dos propósitos definidos inicialmente. Outro respondente destacou que as etapas mais importantes da governança proposta são a *Implementação e Administração da Governança* por perceber que a mudança cultural em relação ao tratamento dos

dados será consequência da execução bem sucedida dessas etapas. Um terceiro respondente acrescentou que considera todas as etapas importantes, e que sem as etapas: *Iniciativas Estratégicas, Visão e Objetivos e Implementação da Governança*, será muito difícil começar um movimento de mudança na organização.

45% acham que a etapa da proposta de Governança de Dados baseada no TOGAF-ADM mais difícil de ser implementada nas empresas é a primeira, ou seja, *Iniciativas Estratégicas*. Em segundo lugar em nível de dificuldade de implementação, 35% do grupo apontou as etapas: *Processos de Negócio, Sistemas de Informação e Implementação da Governança*. Em terceiro lugar em nível de dificuldade de implementação, cerca de 25% do grupo considerou a etapa: *Administração da Governança* como a mais difícil. Em quarto lugar em nível de dificuldade de implementação, somente cerca de 10% do grupo apontou as etapas: *Visão e Objetivos, Tecnologia e Padronização e Administração de Requisitos*.

Um respondente destacou que conscientizar as pessoas sobre a importância dos dados, da qualidade desses dados e rever os processos de negócio da organização é complicado, uma vez que irá provocar mudanças no comportamento das pessoas e também na forma elas tratam a informação, pois considera que ainda é muito forte na Empresa X a cultura dos “feudos” em que a retenção da informação é tratada como ferramenta de poder. Outro acrescentou que rever os processos de negócio existentes causa muita resistência por parte dos envolvidos, seja pela comodidade existente ou pela resistência natural que o ser humano possui à mudança. Outro acrescentou que se houver necessidade de grandes investimentos tecnológicos para a implementação da governança, nem sempre isto está previsto no orçamento das empresas e precisa ser negociado e defendido.

Os questionados foram convidados a destacar quais características que poderiam ser obtidas por uma Governança de Dados baseada em um framework corporativo (Figura 28). Somente um entrevistado comentou que a principal característica estava relacionada à expressão “framework corporativo”, por entender que, o framework, por referir-se a corporação com um todo, deveria estar disponível onde todos possam ter acesso às informações da empresa com qualidade e precisão. 90% acreditam que a *Sistematização dos passos para concretização da governança* é uma das características que a Governança de Dados pode trazer se for baseada em um framework corporativo. As características: *maior facilidade na implantação da governança* e o *melhor alinhamento entre a área de Tecnologia da Informação e as áreas de negócio* ficaram

empatadas em segundo lugar na opinião de aproximadamente 80% dos respondentes. Cerca de 65% dos respondentes apontaram, em terceiro lugar, que a *melhor identificação dos componentes e recursos necessários* é a característica que eles acham que a Governança de Dados pode trazer se for apoiada em um framework corporativo. As demais opções não foram destacadas por nenhum respondente.

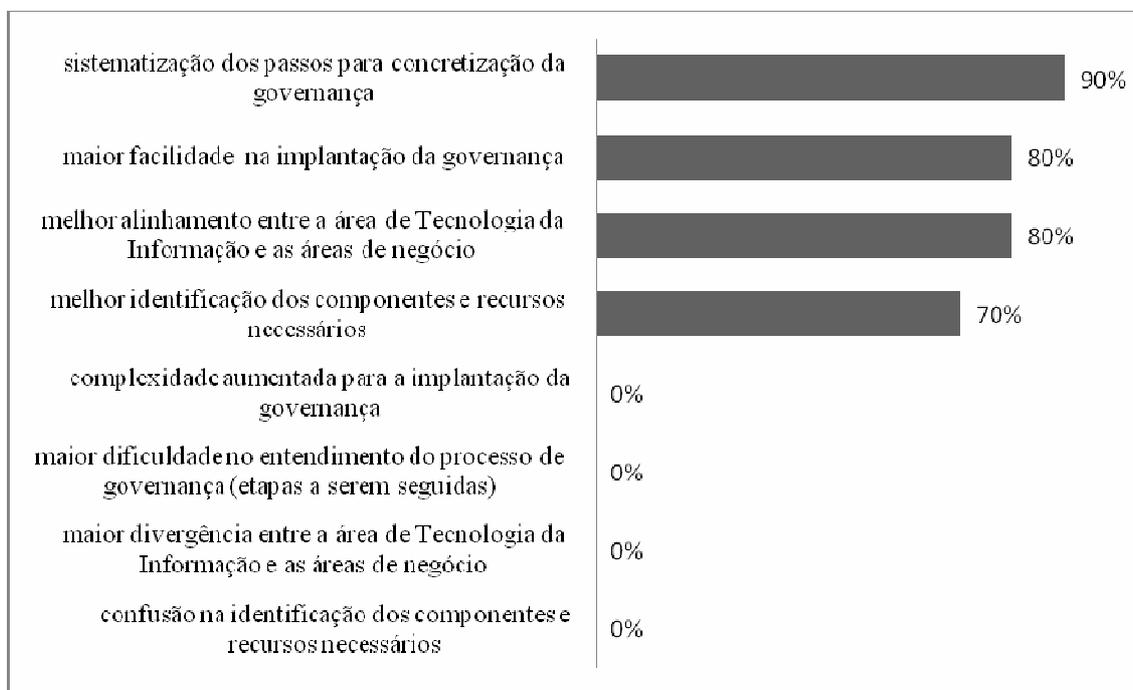


Figura 28 - Características relevantes da Governança de Dados baseada em um Framework Corporativo

Percebe-se que as pessoas não têm noção da quantidade de questões que devam ser consideradas e que são importantes na Governança de Dados. Normalmente, a idéia inicial é de que trata-se de algo relativamente fácil de implementar e que um conjunto de atividades parece ser algo factível, mesmo sem uma metodologia para isto. Mas quando as pessoas são apresentadas a uma metodologia de Governança de Dados, observa-se que sem tal sequência de passos seria muito difícil de obter bons resultados na Governança de Dados.

Em seguida, todos foram convidados a identificar as características e diferenças entre as abordagens de Governança de Dados e Qualidade de Dados (Figura 29). A partir de suas respostas, observa-se as seguintes opiniões: primeiramente todos foram unânimes em afirmar que não acreditam que as abordagens de Governança de Dados nem a de Qualidade de Dados sejam utópicas. 90% dos respondentes acham que as duas abordagens são demoradas e trabalhosas. 70% consideram que tanto a Governança de

Dados quanto a Qualidade de Dados são custosas, enquanto que 25% não consideram nenhuma das duas como custosa.

Enquanto 70% consideram que ambas as abordagens são completas, somente 20% acham que a Governança de Dados é a abordagem mais completa. Da mesma forma que 70% pensam que ambas são indispensáveis, enquanto que 25% acham que somente a Governança de Dados é indispensável.

Quanto à abordagem ser baseada no negócio, 25% acham que ambas (GD e QD) o são, enquanto que 70% consideram que somente a Governança de Dados é baseada no negócio. Isso permite observar como Governança de Dados se encontra mais próxima aos negócios. Dos questionados, 55% consideram que tanto a Governança de Dados quanto a Qualidade de Dados são baseadas em TI, enquanto 35% consideram que somente a Qualidade de Dados é que se baseia em TI. Isto se justifica uma vez que o termo basear pode ser entendido no sentido de apoiar ou de fundamentar. A TI certamente apóia iniciativas de Governança de Dados e de Qualidade de Dados, embora o segundo tenha seu enfoque maior em TI.

E quanto à abordagem ser técnica e sistemática, observa-se que 45% acham que a abordagem em Qualidade de Dados é técnica e sistemática enquanto que 45% acham que ambas as abordagens são técnica e sistemática. Este resultado corrobora com esta visão mais técnica de Qualidade de Dados em relação à Governança de Dados.

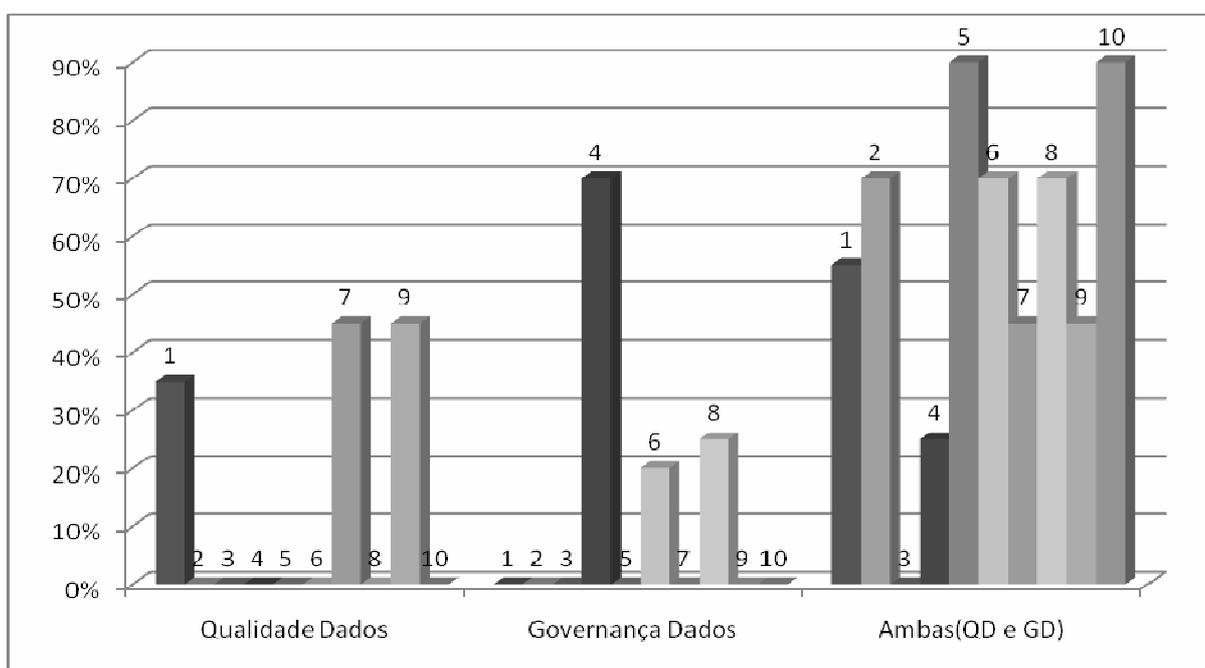


Figura 29 - Características das abordagens para Qualidade de Dados, Governança de Dados e ambas

Legenda:

| | | | |
|---|----------------------|----|-----------------|
| 1 | - Baseada na TI | 6 | - Completa |
| 2 | - Custosa | 7 | - Técnica |
| 3 | - Utópica | 8 | - Indispensável |
| 4 | - Baseada no negócio | 9 | - Sistemática |
| 5 | - Demorada | 10 | - Trabalhosa |

A seguir, os questionados foram estimulados a apontar os pontos fortes da Governança de Dados proposta. Inicialmente apontaram como ponto forte a apresentação de uma metodologia de forma a se buscar e garantir qualidade dos dados, além de possibilitar a identificação e priorização dos problemas a serem resolvidos. A proposta de um ciclo contínuo apresentada no framework também foi apontada como um ponto forte. Também apontaram a clareza com que se obtém melhoria nas avaliações e tomada de decisão bem como na qualidade das informações. A conscientização do papel das áreas responsáveis pela informação e seu tratamento também foi um ponto forte relatado. E também destacaram que o framework mostra a necessidade de tratamento dos dados para o cumprimento das normas regulamentares.

Outros destacaram a necessidade da divulgação da importância dos dados dos clientes para todos os colaboradores. A importância da gestão dos dados através de batimento de informações, demonstrativo de inconsistências e existência de auditoria do processo, promovendo assim maior sensibilização de toda empresa ao cuidado com os dados. Alguns comentaram que seria importante implementar uma cultura de consciência da relevância da manutenção dos dados, a partir da implantação da governança, podendo ou devendo a direção da organização autorizar recursos para essa validação.

Outros respondentes realçaram como ponto forte o tratamento dos dados como um bem da empresa, um ativo precioso, por esta razão, a importância em definir estratégias e mecanismos para manter este ativo sempre valorizado, disponível e com qualidade para ser usado por toda organização. Outros ainda enfatizaram que a divulgação do conceito de Governança de Dados para toda e qualquer pessoa que possa utilizá-la, bem como a clara definição das responsabilidades sobre os dados.

Uma das questões do questionário solicitava que o entrevistado apontasse os pontos fracos da Governança de Dados proposta. Cerca de 30% dos respondentes não

conseguiram apontar nenhum ponto fraco na proposta apresentada. Dos restantes, o principal ponto apontado como fraco foi: “se não houver conscientização das áreas de negócio e TI e não houver patrocínio da alta administração a governança não se realiza efetivamente”. Isso só vem reforçar a importância da Governança de Dados estar contribuindo para a união das duas áreas em prol do sucesso deste trabalho, bem como buscando o patrocínio e o envolvimento da alta administração logo na primeira fase (neste caso, na *Iniciativa Estratégica*). Outro apontou a ausência na apresentação de uma lista contendo os critérios de qualidade mais utilizados para dados. No estudo de caso só foi apresentado o que foi utilizado, e este respondente ficou com a sensação de que poderia estar faltando algo pelo desconhecimento de todas as possíveis opções.

Uma profissional da área de TI fez questão de destacar que apesar de ter o entendimento claro de que a governança é uma ação contínua e de não se tratar simplesmente de um projeto, ela não teria certeza se as empresas conseguiriam imprimir os esforços necessários sem que existisse a conotação de projeto. Esta abordagem parece coerente com a realidade vivida na Empresa X, onde se tenta trabalhar por projeto já há algum tempo, porém ainda com muita dificuldade, e no caso da área de TI, esta é constantemente “atropelada” por novas e sempre urgentes requisições que acabam por prejudicar o andamento de alguns projetos. Sendo assim, como se poderia pensar em êxito em uma ação contínua de governança, se a empresa não conseguiu ainda absorver essa forma de trabalho. Isso só vem reforçar que o trabalho de divulgação e conscientização organizacional são extremamente importantes na Governança de Dados, possibilitando a mudança cultural na organização. Outro ainda acrescentou que considera a dificuldade de conscientização corporativa como um ponto fraco e que como um projeto de longo prazo existe a exigência de forte alocação de recursos. Outro destacou que é necessário demonstrar a relação *Custo versus Benefício* com base no índice de exatidão dos dados, em face dos processos existentes, e áreas afins que atuam nessa atividade e o quanto isso implica diretamente nas despesas administrativas.

Em seguida, os questionados foram convidados a relacionar sua convicção em relação ao que toda proposta de Governança de Dados deveria ter. Eles apontaram que primeiramente o apoio integral e patrocínio da alta administração e a atenção e foco na importância do assunto por parte do corpo executivo. Um patrocinador forte, pois dessa forma seria mais fácil conscientizar toda organização; sem apoio da alta administração é muito difícil ter ou implementar uma governança. Outros apontaram a importância da existência de uma equipe multidisciplinar que atuaria permanentemente através de um

comitê de acompanhamento com representantes de todas as áreas envolvidas. Outros ainda destacaram que é primordial a divulgação dentro da empresa, ou seja, uma etapa de divulgação cultural chamando atenção para importância da informação confiável. Outros ainda destacaram que a atualização tecnológica permanente, sempre com as ferramentas de mercado mais adequadas e manutenção de indicadores que apontem as melhorias no tratamento dos dados. E por último, a existência de um framework corporativo para melhorar o alinhamento entre as áreas de TI e de negócio, conjugando o apoio necessário e facilitando a implementação da governança.

Se a organização optar por seguir a proposta de governança baseada no TOGAF-ADM, o grupo de profissionais da Empresa X que participou desta avaliação apoiaria incondicionalmente. Um dos questionados acrescentou ainda que o tema abordado aqui tem a pretensão de tratar os dados como um bem que merece atenção e qualidade nas organizações. Lembrou que as organizações só foram despertar para a necessidade de garantir a segurança das informações há um tempo não muito distante, e que percebe como um processo evolutivo, onde o próximo passo seria implementar a Governança de Dados, garantindo assim mais confiabilidade e consistência nas informações.

6. Conclusão

Neste trabalho discutiu-se a importância dos dados para as organizações, que devem se empenhar em tratá-los com a devida qualidade. Questões relacionadas a propostas e metodologias envolvendo qualidade e Governança de Dados foram apresentadas. Foi feita uma pesquisa exploratória junto a profissionais de TI (principalmente os envolvidos em dados de apoio à decisão) para avaliar como estão lidando com a questão da qualidade dos dados. Também se quis avaliar a necessidade de apoiar as equipes de TI nas atividades relacionadas à gerência da qualidade, incluindo a explicitação de critérios de qualidade nos dados, especificamente em *Data Warehouse* e/ou *Data Marts*.

Na tentativa de analisar as respostas da pesquisa exploratória desenvolvida através de um questionário on-line que tentava medir a reação dos profissionais de TI quanto à qualidade de dados, evidenciou-se claramente que este assunto deve ser tratado pela organização como um programa de Governança de Dados, de ação contínua e que envolva a alta direção, os gestores das áreas de negócio, os profissionais da TI envolvidos com administração e qualidade dos dados.

Foi proposta uma abordagem de Governança de Dados baseada no TOGAF-ADM, possibilitando uma visão corporativa no tratamento dos dados. Esta proposta apresenta uma metodologia de desenvolvimento da Governança de Dados de modo a permitir melhor alinhamento entre TI e Negócios. Um estudo de caso foi realizado no contexto de uma empresa de previdência complementar e os resultados indicaram a importância e a viabilidade da proposta.

Na próxima seção, apresenta-se uma comparação com outras propostas de Governança de Dados (Seção 6.1), bem como são ressaltadas as principais contribuições (Seção 6.2) e indicados alguns trabalhos futuros (Seção 6.3).

6.1 Comparação com outras propostas de Governança de Dados

Segundo o PMI (*Project Management Institute*), um programa pode ser conceituado como um conjunto de Projetos que se relacionam de alguma forma e que se tratados em conjunto tendem a trazer benefícios globais, onde projetos gerenciados de forma individual não obteriam. Os frameworks descritos no Capítulo 2 tratam a Governança de Dados como um programa, faltando a visão empresarial que uma arquitetura empresarial proporciona. A proposta da IBM (IBM DATA GOVERNANCE COUNCIL MATURITY MODEL, 2007) enumera elementos de governança, dividindo em 11 disciplinas distribuídas em 4 grupos. Já a Oracle (HSU, 2009) e o DGI (GWEN, 2009) tratam como programa, o primeiro dividido em 6 partes e o segundo contemplado por 10 componentes divididos em 3 grupos (Anexo III). O TOGAF-ADM provê uma abordagem global de *design*, planejamento, implementação e governança de uma arquitetura de informação corporativa, possibilitando que a Governança de Dados seja tratada como uma ação contínua e esse foi o principal motivo de ter sido utilizada neste trabalho.

É possível observar que o alinhamento dos objetivos estratégicos com os de Governança de Dados não é mencionado pela IBM, enquanto é mencionado pelas demais propostas. Porém a abordagem de que esses objetivos estratégicos devam ser perseguidos pela equipe de Governança de Dados somente é contemplada pela Oracle e por este trabalho.

Já o alinhamento da estratégia de qualidade de dados caminhando junto com a estratégia do negócio não é mencionado pelo DGI, enquanto é contemplado pelos demais. Neste trabalho, encontra-se a definição de que as estratégias de qualidade de dados sejam tratadas no planejamento estratégico, inclusive definindo medidas de desempenho e administração, apesar da Oracle também defender a existência de uma definição da estratégia da informação na organização.

Pode-se observar que não existe um consenso quanto às responsabilidades e papéis, porém todas as propostas apontam para a necessidade de se definir grupos de pessoas que devem estar envolvidas na Governança de Dados. A única que não aborda a qualificação dos papéis é a IBM. Nesta dissertação, é sugerido um comitê de governança que possua uma formação dinâmica, proporcionando às pessoas a atuarem na gestão estratégica e na operacional, quando necessário. Também foi acrescentado o papel da auditoria e a equipe de avaliação de riscos e controle, para garantir a

conformidade e a continuidade do processo de avaliação e comprovação da governança na organização.

Pode-se ainda destacar que os pontos comuns perseguidos por todas as propostas são a existência de definição de políticas de negócios que guiem a coleção e administração de dados, de políticas e processos de Governança de Dados, e de políticas e procedimentos para padronização de dados.

Este trabalho destaca também a necessidade de se manter um manual de políticas e procedimentos de Governança de Dados, bem como um manual para orientar a modelagem de dados. Outra distinção feita neste trabalho em relação aos demais diz respeito à identificação das regulamentações existentes e que estão sendo atendidas pela Governança de Dados, apesar de que todas as propostas, com exceção da Oracle, possuem ações com o objetivo de mitigar riscos operacionais relacionados à segurança de TI. Este trabalho aborda a importância da existência de um plano de contingência, de continuidade de negócios (com Administração de Crise, Resposta a Incidentes, Contingência Operacional, Recuperação de Desastre, e Testes e Simulações), de continuidade dos processos críticos de TI da empresa e de continuidade de negócios da rede de telecomunicações, todas características devidamente contempladas pela proposta de Governança de Dados.

Exceto o DGI, as demais propostas contemplam a continuidade de negócios para banco de dados, aquisição e retenção de dados externos, bem como definição de procedimentos de infra-estrutura e segurança física e identificação de problemas existentes na qualidade da base de dados e correção dos dados.

No framework da IBM não existe a definição clara da importância da divulgação das ações de governança para todos os membros da organização, fundamental para o acompanhamento, entendimento e aculturação de todos. A comunicação ineficaz dificulta as relações nas organizações. A proposta da Oracle, bem como a do DGI, também não aborda treinamento do usuário, nem aculturação da organização para uma efetiva Governança de Dados. A cultura organizacional deve direcionar a forma de perceber, pensar e sentir a realidade dentro da organização, da mesma forma que serve como um modelo para lidar com os problemas de adaptação e interação. Por este motivo, é importante a cultura organizacional fazer parte da Governança de Dados.

6.2 Contribuições da Pesquisa

Este trabalho destaca a crescente percepção nas organizações de que uma estratégia bem definida para tratamento dos dados é fundamental para o seu sucesso. E para tal, surge uma ação multidisciplinar de Governança de Dados, que promova o alinhamento entre as áreas de TI e Negócio, objetivando a melhoria contínua da governança e qualidade dos dados. Propõe um *framework* de Governança de Dados baseada no *TOGAF-ADM*, utilizando uma sequência iterativa de passos para criação e desenvolvimento de uma arquitetura empresarial completa, capaz de elucidar todos os passos necessários para uma efetiva Governança de Dados em uma organização.

Cada etapa encontra-se detalhada, mostrando uma sequência de execução e possibilitando uma visão única a todos os envolvidos na governança, colocando-os no mesmo nível de importância e responsabilidade. Este *framework* não se encontra relacionado a nenhum fornecedor de software, logo não se corre o risco de ficar restrito a uma organização de consultoria específica com sua adoção. O estudo de caso aplicado em uma empresa permitiu uma avaliação dessa proposta junto aos profissionais de Negócio e TI. Além disso, este trabalho buscou elucidar outras propostas de Governança de Dados existentes na literatura, bem como resumir os critérios de qualidade de dados e metodologias relacionadas à qualidade de dados também discutidas na literatura.

6.3 Perspectivas Futuras

O estudo de caso apresentado nesta pesquisa teve um escopo bem reduzido, executando-se as fases do *framework* em uma empresa de previdência e somente uma única vez. Sugere-se que esta proposta de governança deva ser aplicada em outras organizações de ramos de negócio, tamanho e características diferentes do que foi apresentado neste trabalho. E também que esta proposta seja aplicada executando-se algumas vezes os ciclos do *framework* para possibilitar uma melhor análise dos resultados. Com isso, este *framework* poderá evoluir de forma que estejam definidas as melhores práticas para cada tipo de organização em cada fase de execução do mesmo, facilitando ainda mais seu entendimento e utilização.

Como foi apontado na avaliação por um gerente de negócio como uma necessidade, sugere-se que o *framework* poderá evoluir para possuir também

mecanismos, tanto quantitativos como qualitativos, de avaliação da conformidade (*compliance*) da implementação da Governança de Dados.

Assim acredita-se caminhar na direção da preservação dos dados existentes na organização, tratando-os como um ativo valioso e precioso que é.

Referências

- ABRAMS, R. M., 1991, *The successful business plan: secrets & strategies*. New York: Oasis.
- AMARAL, G., 2003, *Aquaware : Um Ambiente de Suporte à Qualidade de Dados em Data Warehouse*. Tese de Mestrado. Instituto de Matemática – Universidade Federal do Rio de Janeiro. Rio de Janeiro, Brasil.
- ANDRADE, M. M., 2001, *Como Preparar Trabalhos para Cursos de Pós Graduação*. 4 ed. São Paulo: Atlas.
- BATINI, C., CAPPIELLO, C., FRANCALANCI, C., e MAURINO, A., 2009, “Methodologies for data quality assessment and improvement”. *ACM Comput. Surv.* 41, 3, Article 16, 52 pp.
- BASILI, V.R., CALDIERA, G., ROMBACH, H.D., 1994, *Goal Question Metric paradigm*. Encyclopedia of Software Engineering. John Wiley & Sons. 2 v.
- BATINI, C. and SCANNAPIECO, M., 2006, *Data Quality: Concepts, Methodologies and Techniques*. Springer Verlag.
- BATINI, C., CABITZA, F., CAPPIELLO, C., e FRANCALANCI, C., 2008, “A comprehensive data quality methodology for Web and structured data”. *Int. J. Innov. Comput. Appl.* 1, 3, 205–218.
- BELCHIOR, A.D., 1992, *Controle da Qualidade de Software Financeiro*. Tese de Mestrado, COPPE/UFRJ; Rio de Janeiro
- BHALLA, S. K., 1987, *The effective management of technology: a challenge for corporations*. New York: Battelle Press.
- BPM-ADVISOR, 2010, “Arquitetura Empresarial”. Disponível em: <http://www.bpm-advisor.com.br/aell1sn01.htm>. Acesso em: março/2010.
- CAPPIELLO, C., FRANCALANCI, C., e PERNICI, B., 2003, “Preserving Web sites: A data quality approach”. In Proceedings of the 7th International Conference on Information Quality (ICIQ).

- CAMPOS, A. L. de S., 2005, *Indicadores de Qualidade de Dados em Data Warehouse*. Monografia de conclusão de curso. UFJF, Juiz de Fora, 51.
- CAPPIELLO, C., FRANCALANCI, C., PERNICI, B., PLEBANI, P., e SCANNAPIECO, M., 2003, “Data quality assurance in cooperative information systems: a multi-dimension certificate”. In Proceedings of the ICDT International Workshop on Data Quality in Cooperative Information Systems (DQCIS).
- COLAÇO, M. Júnior, 2004, *Projetando sistemas de apoio à decisão baseados em data warehouse*. Rio de Janeiro: Axcel Books.
- CHAUDHURI, S. e DAYAL, U., 1997, “Requirements capture and analysis: a survey of current practice”, *Requirements engineering*, pp. 75-88.
- CHENGALUR-S., I. N., BALLOU, D. P., e PAZER, H. L., 1999, “The impact of data quality information on decision making: An exploratory analysis”. *IEEE Trans. Knowl. Data Eng.* 11, 6, 853–864.
- CLEMENTE, F. *apud* GIL, A. C. (2007). “Pesquisa qualitativa, exploratória e fenomenológica: Alguns conceitos básicos”. Disponível em: <http://www.administradores.com.br/informe-se/artigos/pesquisa-qualitativa-exploratoria-e-fenomenologica-alguns-conceitos-basicos/14316/>. Acesso em: abril/2010.
- DAVENPORT, T. H., 2001, *Ecologia da informação*. Tradução: Bernadette Siqueira. Abrão. 3. ed. São Paulo: Futura, 316 p
- DATA WAREHOUSING INSTITUTE, 2006, “Data quality and the bottom line: Achieving business success through a commitment to high quality data”. Disponível em: <http://www.dw-institute.com/>. Acesso em: março/2009.
- D'ANDRÉA, E., 2005, “Cuide de seus Dados”. Disponível em: http://www.itweb.com.br/solutions/banco_de_dados/armazenamento_de_dados/artigo.asp?id1=39254. Acesso em: março/2009.
- DW BRASIL, 2005, “Características de um data warehouse”. Brasília. Disponível em: http://www.dwbrasil.com.br/html/artdw_carac.html. Acesso em: março/2009.
- FILHO, J. T., 2003, “Transformando Dados Corporativos em Inteligência Competitiva”. Disponível em: <http://www.informal.com.br>. Acesso em: março/2009.
- FALORSI, P., PALLARA, S., PAVONE, A., ALESSANDRONI, A., MASSELLA, E., e SCANNAPIECO, M., 2003, “Improving the quality of toponymic data in the

- Italian public administration”. In Proceedings of the ICDT Workshop on Data Quality in Cooperative Information Systems (DQCIS).
- FERREIRA, A.B.H, 1999, *Aurélio Século XXI: O dicionário da língua portuguesa*. 3ed. Rio de Janeiro: Nova Fronteira.
- FISHER, T., 2009, *The Data Asset – How smart Companies Govern their data for Business Success*, John Wiley & Sons, Inc.
- GARTNER GROUP REPORT, 2005, Gartner Press Release, Gartner Website – Media relations. Disponível em: http://www.gartner.com/press_releases/pr2005.html
Acesso em: setembro/2009.
- FRIEDMAN T., DECKER J.E.V., 2009, Gartner FEI Technology Study Reveals Finance Managers' Perspectives on Data Quality, GARTNER GROUP REPORT, ID Number: G00169706
- GIL, A. C., 1996, *Como elaborar projetos de pesquisa*. São Paulo: Atlas.
- International Organization for Standardization, 2000: *ISO 9001:2000: Quality Management Systems – Requirements*. Beuth, Berlin.
- GWEN T., 2009, *The DGI Data Governance Framework*, The Data Governance Institute.
- HSU, V. and O’NEAL K., 2009, *How Technology Enables Data Governance*, Oracle Corporation, World Headquarters, U.S.A. Disponível em: www.oracle.com.
Acesso em: setembro/2009.
- IBM DATA GOVERNANCE COUNCIL MATURITY MODEL, 2007, *The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance*. Disponível em: ibm.com/software/data/information/trust-governance.html. Acesso em: setembro/2009.
- INMON, W. H., 2003. *Enterprise Intelligence – Enabling High Quality In the Data Warehouse / DSS Environment*. Disponível em: http://www.ascentialsoftware.com/cgi-in/dataquality.cgi?URL=WP_Inmon_3.pdf.
Acesso em: outubro/2008.
- INMON, W. H., 2005. *Building the Data Warehouse*, 4th Ed., New York: Wiley.
- ISO 9000, 2005, “Quality management systems – Fundamentals and vocabulary”, International Organization for Standardization.
- JEUSFELD, M., QUIX, C., e JARKE, M., 1998, “Design and analysis of quality information for datawarehouses”. In Proceedings of the 17th International Conference on Conceptual Modeling.

- KIM, Y. J.; KISHORE, R.; SANDERS, G. L., 2005, "From DQ to EQ: understanding data quality in the context of e-business systems." *Communications of the ACM*, v. 48, n. 10, p. 75-81.
- KIMBALL, R. et al., 2000. *Indicators of Quality*. *Intelligent Enterprise*, v.3, n.6. Disponível em: <http://www.intelligententerprise.com/000410/webhouse.shtml>
Acesso em: Novembro/2009.
- KOTTER, J., 2005, *Liderando Mudança*, Editora Campus, ISBN: 8535201939.
- LEE, Y.W., STRONG, D. M., KAHN, B. K., e WANG, R. Y., 2002, "AIMQ: A methodology for information quality assessment". *Inform. Management*. 40, 2, 133-460.
- LINTHICUM D.S., 2003, "Next Generation Application Integration: From simple Information to Web Services", Addison Wesley.
- LIU, L. e CHI, L., 2002, "Evolutionary data quality". In *Proceedings of the 7th International Conference on Information Quality*.
- LONG, J. e SEKO, C., 2005, "A cyclic-hierarchical method for database data-quality evaluation and improvement". In *Advances in Management Information Systems- Information Quality Monograph (AMISIQ) Monograph*, R. Wang, E. Pierce, S. Madnick, and Fisher C.W.
- LOSHIN, D., 2004, "Enterprise Knowledge Management - The Data Quality Approach". *Series in Data Management Systems*, Morgan Kaufmann, cap4.
- McGEE, J. V.; PRUSAK, L., 1994, *Gerenciamento estratégico da informação: aumente a competitividade e a eficiência de sua empresa utilizando a informação*. 6. ed. Rio de Janeiro: Campus, 245 p.
- NBR ISO 8402, 1994 "Gestão da qualidade e garantia da qualidade – Terminologia", Norma Brasileira ISO.
- OLIVEIRA, D. P R., 1999, *Planejamento estratégico*. 14. ed. São Paulo: Atlas.
- OLSON, Jack E., 2003, *Data Quality – The Accuracy Dimension*, Morgan Kaufmann Publishers
- PERNICI, B. and SCANNAPIECO ,M., 2003, "Data quality in Web information systems". *J. Data Semant*. 1, 48-68.
- PINHO, S.F.C., 2001, *Avaliação da Qualidade de Dados pela Não Conformidade*. Dissertação de Mestrado – COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

- PIPINO, L., LEE, Y., e WANG, R., 2002, “Data quality assessment”. *Commun. ACM* 45, 4.
- PROFORMA, 2003, “Sarbanes-Oxley Express - A process-driven Path to Compliance”. Proforma Corporation. Disponível em: <http://www.proformacorp.com>. Acesso em: setembro/2009.
- RAISINGHANI, M., 2004, *Business Intelligence in the Digital Economy*, Hershey PA: The Idea Group
- RAU, K. G., 2004, Effective governance of IT: Design objectives, roles, and Relationships. *Inf. Syst.Manag.* 21, 4, 35–42.
- REDMAN, T.C., 1998, “The Impact of Poor Data Quality on the Typical Enterprises” *Communications of ACM*, 41, 2.
- SCANNAPIECO, M., A. VIRGILLITO, MARCHETTI, M., MECELLA, M., e BALDONI, R., 2004, “The DaQuinCIS architecture: a platform for exchanging and improving data quality in Cooperative Information Systems”. *Inform. Syst.* 29, 7, 551–582.
- SCANNAPIECO, M., PERNICI, B., e PIERCE, E., 2002, “IP-UML: Towards a Methodology for Quality Improvement based on the IP-MAP Framework”. In *Proceedings of the 7th International Conference on Information Quality (ICIQ)*. Boston.
- SCANNAPIECO, M., PERNICI, B., e PIERCE, E., 2005, “IP-UML: A methodology for quality improvement-based on IP-MAP and UML”. In *Information Quality, Advances in Management Information Systems, Information Quality Monograph (AMIS-IQ)*, R. Wang, E. Pierce, S. Madnik, e C. Fisher, Eds.
- SCHEKKERMAN, J., 2008, *Enterprise Architecture Good Practices Guide: How to Manage the Enterprise Architecture Practice, Perfect Bound Softcover(B/W)*, ISBN: 9781425156879
- SHAMS K., FARISHTA M., 2001. “Data Warehousing: Toward knowledge Management”, *Topics in Health Information Management*, v. 21, n. 3, p. 24-32.
- SHANKARANARAYAN, G.; ZIAD, M.; WANG, R., 2003, “Managing data quality in dynamic decision environments: an information product approach”. *Journal of Database Management*, v. 14, n. 4, p. 14-32.
- SU, Y. e JIN, Z., 2004, “A methodology for information quality assessment in the designing and manufacturing processes of mechanical products”. In *Proceedings of the 9th International Conference on Information Quality (ICIQ)*. 447–465.

- THE OPEN GROUP, 2009, TOGAF Version 9: The Open Group Architecture Framework (TOGAF), ISBN: 978-90-8753-230-7, Document Number: G091
- THE OPEN GROUP, 2008, The Open Group Website. Disponível em: www.opengroup.org/architecture/Togaf8-doc/arch/toc-tp4.html. Acesso em: julho/2010.
- VERGARA, S. C., 2004, Projetos e relatórios de pesquisa em Administração. São Paulo: Atlas.
- WAND, Y. e WANG, R., 1996, "Anchoring data quality dimensions in ontological foundations". Comm.ACM 39, 11.
- WANG, R., 1998, "A product perspective on total data quality management". Comm. ACM 41, 2.
- WANG, R. e STRONG, D., 1996, "Beyond accuracy: What data quality means to data consumers". J. Manage.Inform. Syst. 12, 4.
- WIXOM, B.; TODD, P., 2005, "A theoretical integration of user satisfaction and technology acceptance". Information Systems Research, v. 16, n. 1, p. 85-102.
- WIXOM, B.; WATSON, H., 2001, "An empirical investigation of the factors affecting data warehousing success", MIS Quarterly, v. 25, n. 1, p. 17-41.
- XAVIER, M. da P. T. e GOMES, S. B., 1999, *A Informação como Vantagem da Empresa Competitiva*. Developer's Magazine, Rio de Janeiro, p. 26-29.
- ZACHMAN, J.A., 2009, "The Framework for Enterprise Architecture: Background, Description and Utility", Zachman Institute for Framework Advancement (ZIFA). Id.Documento= 810-231-0531
- ZACHMAN, J. A., 2009, "The Zachman Framework Evolution". Disponível em: <http://www.zachmaninternational.com/index.php/ea-articles/100>. Acesso em: março/2010.
- ZACHMAN, J. A., 2009, "The Zachman Framework: The Official Concise Definition". Disponível em: <http://www.zachmaninternational.com/index.php/home-article/13>. Acesso em: março/2010.
- ZACHMAN, J. A., 2009, "The Zachman on the Framework". Disponível em: [http://xpertaml.com/backup/ABS%20Development%20\(Martin\)/Methodologies/ZIFA/ZIFA09.pdf](http://xpertaml.com/backup/ABS%20Development%20(Martin)/Methodologies/ZIFA/ZIFA09.pdf). Acesso em: março/2010.
- ZAMAN, M., 2005, "Business Intelligence: Its Ins and Outs". Disponível em: <http://technologyevaluationcenters.com>. Acesso em: março/2010.

ZIEGLER, P., DITTRITCH, K., 2007, Data Integration — Problems, Approaches, and Perspectives. Em: John Krogstie, Andreas L. Opdahl, and Sjaak Brinkkemper, editors, *Conceptual Modelling in Information Systems Engineering*, pág. 39–58. Springer, Berlin.

ANEXO I

PESQUISA EM QUALIDADE DE DADOS EM DW E/OU DATA MARTS EM EMPRESAS SEDIADAS NO BRASIL

VOCÊ E A ORGANIZAÇÃO QUE VOCÊ TRABALHA:

1. Nome: _____
2. Empresa: _____
3. Cargo que ocupa atualmente: _____
4. Há quanto tempo trabalha na organização ?
 menos de 1 ano 1 ano 2 anos 3 anos 4 anos 5 anos
 6 anos 7 anos 8 anos 9 anos mais de 10 anos
5. Há quanto tempo atua na área de TI?
 menos de 1 ano 1 ano 2 anos 3 anos 4 anos 5 anos
 6 anos 7 anos 8 anos 9 anos mais de 10 anos
6. Há quanto tempo atua na área de DW e/ou Data Mart?
 menos de 1 ano 1 ano 2 anos 3 anos 4 anos 5 anos
 6 anos 7 anos 8 anos 9 anos mais de 10 anos
7. Há quanto tempo foi iniciado o projeto de DW e/ou Data Mart na sua organização?
 menos de 1 ano 1 ano 2 anos 3 anos 4 anos 5 anos
 6 anos 7 anos 8 anos 9 anos mais de 10 anos
8. Há quanto tempo a primeira aplicação OLAP foi disponibilizada aos usuários na sua organização?
 menos de 1 ano 1 ano 2 anos 3 anos 4 anos 5 anos
 6 anos 7 anos 8 anos 9 anos mais de 10 anos
9. Quais publicações (livros, jornais, revistas, sites etc.) e eventos relacionados a Data Warehouse/BI/OLAP você lê/participa e com qual frequência?

| Nome da Publicação/Evento | Frequência | Utilidade |
|---------------------------|------------|-----------|
| | | |
| | | |
| | | |
| | | |

Utilidade: de 0 a 100, onde 0 = sem utilidade e 100 = extremamente útil

Frequencia: Diária, Semanal, Mensal, Bimestral, Trimestral, Anual

10. Quais publicações (livros, jornais, newsletters digitais, blogs, white-papers etc.) e eventos relacionados a Qualidade de Dados você lê/participa e com qual frequência?

| Nome da Publicação/Evento | Frequência | Utilidade |
|---------------------------|------------|-----------|
| | | |
| | | |
| | | |
| | | |
| | | |

Utilidade: de 0 a 100, onde 0 = sem utilidade e 100 = extremamente útil

Frequência: Diária, Semanal, Mensal, Bimestral, Trimestral, Anual

ESTRATÉGIA ORGANIZACIONAL NO TRATAMENTO DOS DADOS:

Na sua organização você considera ou percebe que ...

- Uma ou mais pessoas na administração da sua organização patrocinam projetos que implementam melhoria contínua da qualidade dos dados.
 sempre com frequência raramente nunca não posso afirmar
- Uma ou mais pessoas na sua organização são responsáveis (coordenador/executor) por ações de melhoria na qualidade dos dados transacionais.
 sempre com frequência raramente nunca não posso afirmar
- Pessoas que pertencem às áreas de negócio da organização estão envolvidas com ações de melhoria na qualidade dos dados.
 sempre com frequência raramente nunca não posso afirmar
- A documentação dos modelos de dados é mantida atualizada pela equipe de TI.
 sempre com frequência raramente nunca não posso afirmar
- Existe um processo que avalia a qualidade dos dados existentes na organização.
 sempre com frequência raramente nunca não posso afirmar
- Na sua organização existe um repositório de metadados constando o significado dos dados existentes e que atende plenamente as expectativas dos usuários.
 Discordo Discordo parcialmente Concordo parcialmente Concordo totalmente não posso afirmar
- A utilização de dados com qualidade permitirá que a sua organização seja:

- | | |
|---|--|
| <input type="checkbox"/> mais ágil | <input type="checkbox"/> mais burocrática |
| <input type="checkbox"/> mais competitiva | <input type="checkbox"/> menos competitiva |
| <input type="checkbox"/> mais lucrativa | <input type="checkbox"/> menos lucrativa |
| <input type="checkbox"/> mais eficiente | <input type="checkbox"/> menos eficiente |
| <input type="checkbox"/> mais organizada | <input type="checkbox"/> menos organizada |

DATA WAREHOUSE E/OU DATA MART:

No que se refere ao projeto de DW e/ou Data Mart, você considera ou percebe que ...

1. A missão e os objetivos do Data Warehouse e/ou Data Mart encontram-se bem definidos.
 sempre com frequência raramente nunca não posso afirmar
2. A documentação das fontes de dados do projeto do DW e/ou Data Mart
 não existe
 é feita no desenvolvimento do sistema fonte
 é feita no início do projeto de DW/Data Mart
 é feita na medida em que o projeto de DW/Data Mart necessita
 não posso afirmar
3. Existe um repositório central de metadados que é utilizado para registrar a documentação das fontes de dados utilizadas no DW e/ou Data Mart.
 sempre com frequência raramente nunca não posso afirmar
4. É utilizada uma metodologia que define conceitos e procedimentos para a definição, medição, análise e melhoria da qualidade dos dados
 TDQM (*Total Data Quality Management*) GQM (*Goal-question Metric*)
 ISO 9001 não utilizamos
 Outra(s) _____
5. Os atributos que possuem conflitos de domínios ou necessidade de conversão, são tratados no processo de limpeza, antes de ir para o DW e/ou Data Mart, alimentando um repositório de metadados com a devida identificação desse tratamento.
 sempre com frequência raramente nunca não posso afirmar
6. Atributos inconsistentes são facilmente detectados no processo de ETL e sua identificação mapeada em um repositório de metadados registrando os dados suspeitos de má qualidade.
 sempre com frequência raramente nunca não posso afirmar

7. Fontes de dados com regras de negócio inconsistentes são facilmente detectadas no processo de ETL e sua identificação mapeada em um repositório de metadados registrando os dados suspeitos de má qualidade.
- sempre com frequência raramente nunca não posso afirmar
8. Violação de regras de integridade nas fontes de dados são facilmente detectadas no processo de ETL e sua identificação mapeada em um repositório de metadados registrando os dados suspeitos de má qualidade.
- sempre com frequência raramente nunca não posso afirmar
9. Todo e qualquer processo de tratamento e limpeza dos dados deve acontecer antes da inclusão no DW e/ou Data Mart, garantindo a qualidade dos dados que estão inseridos nesta base.
- sempre com frequência raramente nunca não posso afirmar
10. O processo de tratamento e limpeza dos dados deve gerar correção no sistema de origem dos dados para que posteriormente possa ser carregado no DW e/ou Data Mart.
- sempre com frequência raramente nunca não posso afirmar
11. Na agregação de valores, a regra de sumarização dos dados envolvidos encontra-se identificada e mapeada em um repositório de metadados.
- sempre com frequência raramente nunca não posso afirmar
12. Na sumarização de um determinado dado, o grau no qual este conteúdo é válido, é quantificado e armazenado em um repositório de metadados.
- sempre com frequência raramente nunca não posso afirmar
13. O processo de tratamento e limpeza dos dados garante a qualidade de dados no DW e/ou Data Mart.
- 100% 90 a 99% 80 a 89% 70 a 79% 60 a 69% Menos de 50%
14. Formal ou informalmente, a equipe de DW mantém-se atualizada quanto a real satisfação dos usuários com o conteúdo do DW e/ou Data Mart.
- sempre com frequência raramente nunca não posso afirmar
15. Os usuários do DW e/ou Data Mart manifestam à equipe de TI a necessidade de ter mais informações sobre a qualidade do conteúdo do mesmo.
- sempre com frequência raramente nunca não posso afirmar

16. Mapear critérios de qualidade de dados no DW e/ou Data Mart é uma tarefa extremamente árdua, pois não possuímos metodologia e/ou ferramentas que apóiem este processo.

Discordo Discordo parcialmente Concordo parcialmente Concordo totalmente

ANEXO II

O presente questionário tem como objetivo avaliar a proposta de Governança de Dados aplicada em um estudo de caso na Empresa X. Este fará parte do trabalho de conclusão do Programa de Pós-Graduação em Informática – Mestrado, ministrado pela Universidade Federal do Rio de Janeiro – UNIRIO – Centro de Ciências Exatas e Tecnológicas.

Você e a organização que trabalha:

1. Há quanto tempo trabalha na organização ?

menos de 1 ano 1 a 5 anos 6 a 9 anos 10 a 14 anos 15 a 20 anos mais de 20 anos

2. Que cargo você ocupa atualmente?

3. Qual é sua área de formação?

4. Como você classificaria sua área de atuação?

Negócio Sistemas Suporte Dados Outro(s)(especificar):_____

No que se refere à Proposta de Governança de Dados apresentada, você considera ou percebe que ...

5. A participação, engajamento e direcionamento da alta administração no apoio às tarefas relacionadas à Governança de Dados...

é 100% vital para o seu êxito é fundamental é importante, mas não essencial

é desejável não é pertinente

Comentários:

6. A visão através de um framework corporativo apresentando a Governança de Dados aos CEOs da empresa, poderá proporcionar...

uma maior adesão e receptividade, uma vez que a governança está explicitada

uma maior conscientização de como a questão deve ser tratada

uma discussão interna motivadora sobre a questão

algum interesse sobre o assunto

confusão sobre como a questão deve ser tratada

uma menor adesão e receptividade, uma vez que aumentará a carga de trabalho dos envolvidos

conflitos de interesse e poder na organização

desinteresse sobre o assunto

Comentários:

7. Qual (ou quais) das etapas da proposta de governança segundo o TOGAF-ADM você achou mais complicada de entender?

- | | | |
|--|--|--|
| <input type="checkbox"/> iniciativas estratégicas | <input type="checkbox"/> visão e objetivos | <input type="checkbox"/> processos de negócio |
| <input type="checkbox"/> sistemas de informação | <input type="checkbox"/> tecnologia e padronização | <input type="checkbox"/> oportunidades e soluções |
| <input type="checkbox"/> plano de implementação | <input type="checkbox"/> implementação da governança | <input type="checkbox"/> administração da governança |
| <input type="checkbox"/> administração de requisitos | | |

Explique:

8. Qual (ou quais) dos passos da proposta de governança segundo o TOGAF-ADM você achou mais importante?

- | | | |
|--|--|--|
| <input type="checkbox"/> iniciativas estratégicas | <input type="checkbox"/> visão e objetivos | <input type="checkbox"/> processos de negócio |
| <input type="checkbox"/> sistemas de informação | <input type="checkbox"/> tecnologia e padronização | <input type="checkbox"/> oportunidades e soluções |
| <input type="checkbox"/> plano de implementação | <input type="checkbox"/> implementação da governança | <input type="checkbox"/> administração da governança |
| <input type="checkbox"/> administração de requisitos | | |

Explique:

9. Qual (ou quais) dos passos da proposta de governança segundo o TOGAF-ADM você considera ser mais difícil de implementar nas empresas em geral?

- | | | |
|--|--|--|
| <input type="checkbox"/> iniciativas estratégicas | <input type="checkbox"/> visão e objetivos | <input type="checkbox"/> processos de negócio |
| <input type="checkbox"/> sistemas de informação | <input type="checkbox"/> tecnologia e padronização | <input type="checkbox"/> oportunidades e soluções |
| <input type="checkbox"/> plano de implementação | <input type="checkbox"/> implementação da governança | <input type="checkbox"/> administração da governança |
| <input type="checkbox"/> administração de requisitos | | |

Explique:

10. Que características você acha que a Governança de Dados pode trazer se for baseada em um framework corporativo?

- maior facilidade na implantação da governança
- complexidade aumentada para a implantação da governança
- sistematização dos passos para concretização da governança
- melhor alinhamento entre a área de Tecnologia da Informação e as áreas de negócio
- maior dificuldade no entendimento do processo de governança (etapas a serem seguidas)

- maior divergência entre a área de Tecnologia da Informação e as áreas de negócio
- melhor identificação dos componentes e recursos necessários
- confusão na identificação dos componentes e recursos necessários

Comentários:

11. Considerando as diferenças entre abordagens de “Governança de Dados - GD” e “qualidade de dados - QD”, assinale com GD, QD, A (ambas), N (nenhuma) o que mais se identifica com cada característica:

- baseada na TI baseada no negócio técnica sistemática
- custosa demorada indispensável trabalhosa
- utópica completa

Acrescente outras características que associa a cada um destes aspectos e que julgue conveniente.

Qualidade de Dados (QD):

Governança de Dados (GD):

QD e GD (Ambas):

12. Destaco como pontos fortes da Governança de Dados apresentada ...

13. Destaco como pontos fracos da Governança de Dados apresentada ...

14. Tenho convicção que toda proposta de Governança de Dados deveria ter ...

15. Se a sua empresa optasse por seguir a proposta de governança segundo o TOGAF-ADM qual seria a sua atitude:

- apoio incondicional
- colaboraria sem entusiasmo
- não estaria interessado em participar
- defenderia outra abordagem
- consideraria melhor não utilizar qualquer abordagem de governança
- apoio com reservas. Quais? _____

Explique:

Outros comentários e sugestões:

ANEXO III

Quadro comparativo entre as principais propostas de Governança de Dados encontradas na literatura e a que é apresentada nesta dissertação.

| Característica considerada | IBM | Oracle | DGI | Este trabalho |
|---|--|---|--|---|
| Formato da Proposta apresentada | Elementos de GD: 11 disciplinas, agrupadas em 4 grupos | Programa contemplando 6 partes | Programa de governança e administração de dados composto por 10 componentes, divididos em 3 grupos | Ação continua promovida pela implementação de um framework de arquitetura empresarial, com 10 passos |
| Estratégia de qualidade de dados caminhando junto com a estratégia do negócio | Contemplado | Contemplado | Não mencionado | Contemplado |
| Objetivos estratégicos perseguidos pela equipe de Governança de Dados | Não mencionado | Contemplado | Não mencionado | Contemplado |
| Alinhamento de objetivos estratégicos com os de Governança de Dados | Não mencionado | Contemplado | Contemplado | Contemplado |
| Estratégias de qualidade de dados no planejamento estratégico, incluindo medidas de performance e administração | Não mencionado | Não mencionado | Não mencionado | Contemplado |
| Definição da estratégia da informação | Não mencionado | Contemplado | Não mencionado | Contemplado |
| Definição de papéis e responsabilidades de Governança de Dados | Mencionado, porém não detalhado | Comitê Líder e um Comitê Operacional de Governança de Dados | <i>Stakeholders</i> de dados, Escritório de Governança de Dados (DGO) e Administrador de Dados. | Comitê de política e gestão de dados, Auditoria interna, Avaliação de riscos e conformidade, <i>Stakeholders</i> de Dados |

| Característica considerada | IBM | Oracle | DGI | Este trabalho |
|---|----------------|----------------|----------------|----------------------|
| Qualificação dos papéis | Não mencionado | Contemplado | Contemplado | Contemplado |
| Definição de políticas de negócios que guiem a coleção e administração de dados | Contemplado | Contemplado | Contemplado | Contemplado |
| Manual de políticas e procedimentos da Governança de Dados | Não mencionado | Não mencionado | Não mencionado | Contemplado |
| Políticas e processos de Governança de Dados | Contemplado | Contemplado | Contemplado | Contemplado |
| Políticas e procedimentos para padronização de dados | Contemplado | Contemplado | Contemplado | Contemplado |
| Políticas e procedimentos para modelagem de dados | Não mencionado | Não mencionado | Não mencionado | Contemplado |
| Políticas e procedimentos para unificação e limpeza de dados | Contemplado | Não mencionado | Não mencionado | Contemplado |
| Estratégia de comunicação e frequência com que as atualizações, métricas e realizações do programa de governança serão divulgadas | Não mencionado | Contemplado | Contemplado | Contemplado |
| Escopo, restrições, expectativas e validação o contexto da governança | Não mencionado | Não mencionado | Contemplado | Contemplado |
| Visão de Governança de Dados | Não mencionado | Contemplado | Contemplado | Contemplado |
| Objetivos de governança de dados | Contemplado | Contemplado | Contemplado | Contemplado |
| Questões para fundamentação estratégica | Não mencionado | Não mencionado | Contemplado | Contemplado |
| Identificação de quais regulamentações devem ser atendidas | Não mencionado | Não mencionado | Não mencionado | Contemplado |
| Questões relativas ao cumprimento de leis regulatórias | Não mencionado | Não mencionado | Contemplado | Contemplado |

| Característica considerada | IBM | Oracle | DGI | Este trabalho |
|--|----------------|----------------|----------------|----------------------|
| Ações com o objetivo de mitigar os riscos operacionais relacionados à segurança de Tecnologia da Informação | Contemplado | Não mencionado | Contemplado | Contemplado |
| Definição de um plano de contingência | Não mencionado | Não mencionado | Não mencionado | Contemplado |
| Definição de um plano de continuidade de negócios (contendo Administração de Crise, Resposta a Incidentes, Contingência Operacional, Recuperação de Desastre, e Testes e Simulações) | Não mencionado | Não mencionado | Não mencionado | Contemplado |
| Continuidade de negócios dos processos críticos de TI da empresa | Não mencionado | Não mencionado | Não mencionado | Contemplado |
| Continuidade de negócios da rede de telecomunicações | Não mencionado | Não mencionado | Não mencionado | Contemplado |
| Continuidade de negócios para os bancos de dados, aquisição e retenção de dados externos | Contemplado | Contemplado | Não mencionado | Contemplado |
| Definição de procedimentos de infraestrutura e segurança física | Contemplado | Contemplado | Não mencionado | Contemplado |
| Identificação de problemas existentes na qualidade da base de dados | Contemplado | Contemplado | Não mencionado | Contemplado |
| Definição e revisão dos processos de negócio sob a ótica do comportamento da informação. | Contemplado | Não mencionado | Contemplado | Contemplado |

| Característica considerada | IBM | Oracle | DGI | Este trabalho |
|---|----------------|----------------|----------------|----------------------|
| Correção dos dados | Contemplado | Contemplado | Não mencionado | Contemplado |
| Estabelecimento de regras de sincronização para evitar novos problemas de qualidade dados | Não mencionado | Não mencionado | Não mencionado | Contemplado |
| Cultura e comportamento em relação a informação | Não mencionado | Não mencionado | Não mencionado | Contemplado |
| Treinamento e desenvolvimento de habilidades | Não mencionado | Não mencionado | Não mencionado | Contemplado |
| Administração de novos requisitos de informação | Não mencionado | Não mencionado | Não mencionado | Contemplado |

ANEXO IV

Descrição resumida de propostas de frameworks de arquitetura empresarial encontradas na literatura:

A-IV.1 C4ISR

A sigla significa: *Command, Control, Computers, Communications (C4), Intelligence, Sureveillance e Reconnaissance (ISR)*. Este framework promove uma orientação arquitetural compreensiva para tudo o que diz respeito ao Departamento de Defesa dos Estados Unidos (DoD), para assegurar interoperabilidade e custo efetivo dos sistemas militares. Este surgiu como sucessor do TAFIM (*Technical Architecture Framework for Information Management*) que foi oficialmente retirado em janeiro de 2000.

O Framework arquitetural C4ISR está sob constante revisão para generalizá-lo atendendo a todas as áreas funcionais do DoD. Ele também está sendo usado em outras áreas do governo além do setor de Defesa Norte Americano. O Framework arquitetural C4ISR destina-se a garantir que as descrições da arquitetura desenvolvida por vários comandos, serviços e organismos no âmbito do DoD são inter-relacionáveis tecnicamente entre e dentro de cada organização operacional, sistemas e visões da arquitetura e são comparáveis e integráveis. Em particular, o Framework arquitetural C4ISR:

- Assegura que as arquiteturas são integráveis em toda a comunidade de Defesa.
- Estabelece ligações ou segmentos que unem visões operacionais, de sistemas e técnicas de uma arquitetura.
- Fornece a base para uma trilha de auditoria que relaciona os sistemas atuais e sistemas postulados de forma a medir a eficácia nas operações e missões.

A-IV.2 CORBA

O *Object Management Architecture (OMA)* do *Object Management Group (OMG)*, muitas vezes, refere-se livremente ao CORBA como uma arquitetura, como uma Arquitetura de aplicação orientada a objeto, centrada no conceito de um *Object Request Broker (ORB)*. O ORB atua como um centro de comutação, localizando objetos, armazenando definições de interface e implementações de objetos e orquestrando mensagens entre objetos em um ambiente heterogêneo distribuído.

Os serviços do CORBA são um conjunto de serviços de objetos, em baixo nível, disponíveis para todos os objetos, abrangendo funções como criação e deleção, nomeação, serviços de segurança, e muitos outros.

As facilidades CORBA são tratadas horizontalmente com funções de alto nível, tais como distribuição de documentos ou de impressão, adequado para uso em vasta variedade de setores do mercado. As facilidades CORBA de domínio vertical, são interfaces de mercado específicas que irão fornecer instalações comuns para aplicações dentro de um segmento específico de mercado.

A-IV.3 EAP

O EAP (*Enterprise Architecture Planning*) foi idealizado por Steven Spewak como um conjunto de métodos para o planejamento do desenvolvimento de Arquiteturas de Informação, Aplicações e Tecnologia. È recomendado que a abordagem seja desenvolvê-los nessa ordem e alinhar os três tipos de arquitetura em relação umas as outras. O objetivo é assegurar que tais arquiteturas formem os modelos que norteiem, sistemas implementáveis que resolvam problemas reais de negócios. Toda metodologia EAP envolve os seguintes passos:

1. Planejamento de Iniciação: definição do escopo, objetivos, funções e responsabilidades e decidindo qual a metodologia a se utilizar, quem deve estar envolvido, e que ferramentas utilizar.
2. Princípios: desenvolvimento dos princípios fundamentais de apoio efetivo à governança da informação e tecnologia. Esses princípios formam a base para a tomada de decisões da arquitetura, aceitação dos resultados e gestão da migração. Elas são baseadas nas melhores práticas da indústria e da finalidade da empresa, visão e valores e são implementadas através de políticas, procedimentos e padrões.
3. Modelagem de Negócios: Modelagem as atividades em curso e as informações utilizadas, identificando oportunidades de negócios e melhorias de processo.
4. Sistemas Atuais e Tecnologia: definir o que está em vigor hoje para sistemas aplicativos e plataformas tecnológicas. Este é um inventário resumindo os sistemas aplicativos, dados e plataforma tecnológica para proporcionar um orientador (uma base) para os planos de migração de longo alcance.

5. Arquitetura de Dados: o desenvolvimento da arquitetura de dados, incluindo a definição das atividades de negócios e objetos de dados necessários para suportar o negócio.
6. Arquitetura de Aplicações: definição dos principais tipos de aplicações necessárias para gerenciar os dados e apoiar as funções empresariais.
7. Arquitetura Tecnológica: a definição das plataformas necessárias para fornecer uma infra-estrutura tecnológica para os aplicativos que gerenciam os dados e apoiar as funções empresariais.
8. Implementação/Planos de Migração: definindo a sequência de implementação das aplicações, um calendário para a implementação, uma análise custo/benefício, e um detalhamento claro para a migração. Recomendações a nível executivo são feitas para a execução do plano e um plano é desenvolvido para o período de transição depois da execução das atividades de planejamento da arquitetura.
9. Planejamento de Conclusão: relatório final e apresentação dos resultados de gestão.

A metodologia EAP, recomenda a sequência: Arquitetura de Negócios, Arquitetura de Dados, Arquitetura de Aplicações, Arquitetura Tecnológica.

A-IV.4 FEA

O Conselho Federal Norte Americano publicou um guia prático de Arquitetura Federativa Organizacional (versão 1.0) em fevereiro de 2001, o FEA (*Federal Enterprise Architecture*), num projeto de cooperação com o *General Accounting Office* (GAO) e o *Office of Management and Budget* (OMB). O objetivo do documento é fornecer para as agências Norte Americanas Federais orientações ao iniciar, desenvolver, usar e manter arquiteturas nas suas organizações. Este guia oferece um processo completo para implantar, implementar e manter um programa de arquitetura corporativa, e descreve os papéis e responsabilidades necessárias para o sucesso de um programa de arquitetura empresarial. As orientações apresentadas no guia prático devem ser adaptadas por cada agência federal de acordo com suas necessidades. Este guia se concentra nos processos de arquitetura corporativa, produtos e funções e responsabilidades. O guia aborda como a arquitetura de processos corporativos cabe dentro do ciclo global de vida de uma empresa, ou seja, descrevendo em detalhes como os processos dizem respeito a arquitetura corporativa de engenharia da empresa, ao

programa de gestão, e dos processos de Planejamento de Capital e Controle de Investimentos (CPIC), conforme resumido na Figura 30:

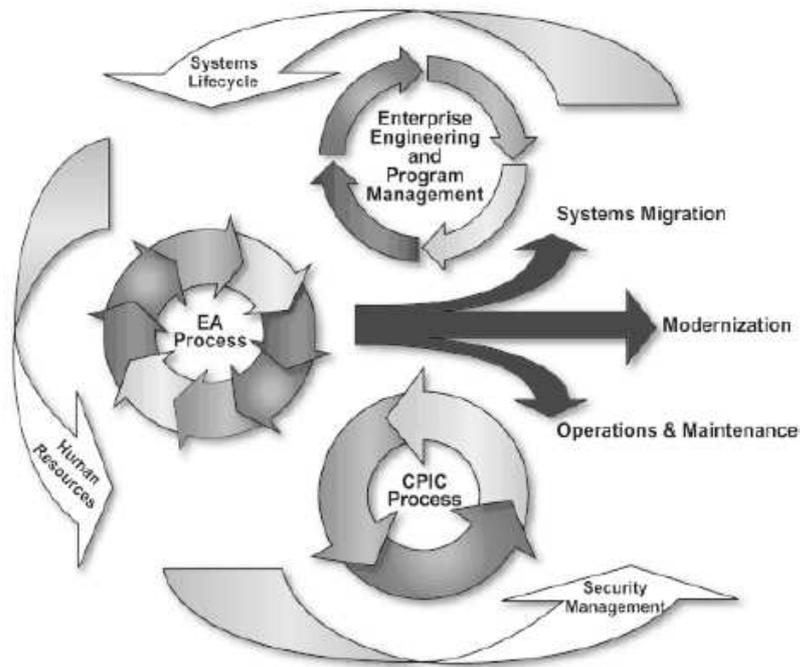


Figura 30 - Ciclo de vida global da organização

No início do seu programa de arquitetura corporativa, cada agência deverá estabelecer o âmbito da sua arquitetura empresarial e formular uma estratégia que inclui a definição de uma visão, objetivos e princípios. A Figura 31 resume a arquitetura de processos corporativos.



Figura 31 - Resumo arquitetura de processos corporativos

Ações de execução e de apoio devem ser estabelecidas e uma equipe de arquitetos formados dentro da organização. A equipe define um processo adaptado às necessidades da agência. A equipe de arquitetura implementa um processo para construir tanto uma

linha de base e o objetivo da Arquitetura. A equipe de arquitetura também prepara um plano de transição para os sistemas, aplicações e práticas de negócio associadas, com base em análises e direcionamento do negócio. Os projetos são selecionados e controlados no CPIC e a engenharia da empresa e o programa de gestão atuam e são guiados pela arquitetura corporativa, e em conformidade com a mesma. Finalmente, a arquitetura é mantida através de modificação contínua para refletir a linha de base da agência e as práticas empresariais, objetivos organizacionais, visões, tecnologia e infraestrutura.

A-IV.5 FEAF

O Conselho Federal Norte Americano publicou um guia prático de Arquitetura, o FEAF (*Federal Enterprise Architecture Framework*) na versão 1.1, em setembro de 1999. O FEAF promove o desenvolvimento compartilhado de processos do Conselho Federal Norte Americano, interoperabilidade e compartilhamento de informações entre suas agências e outras entidades governamentais. O FEAF fornece direção e orientação às agências federais para estruturar uma arquitetura corporativa. O FEAF descreve oito componentes de uma arquitetura empresarial: Diretrizes da Arquitetura, Direção Estratégica, Arquitetura-*Baseline*, Arquitetura-Alvo, Processos de Transição, Segmentos Arquiteturais, Modelos Arquiteturais e Normas.

O FEAF também fornece orientação para o estabelecimento de "segmentos federais", que são áreas de negócio interagências (tais como comércio internacional, bolsas de estudo, prontuários comum) que transcendem as fronteiras de uma agência federal. Esses segmentos da arquitetura federativa coletivamente constituem a Arquitetura Federativa da organização. As partes da FEAF são compostas de Arquitetura de Negócios, Dados, Aplicações e Tecnologia, como mostrado na Figura 32.

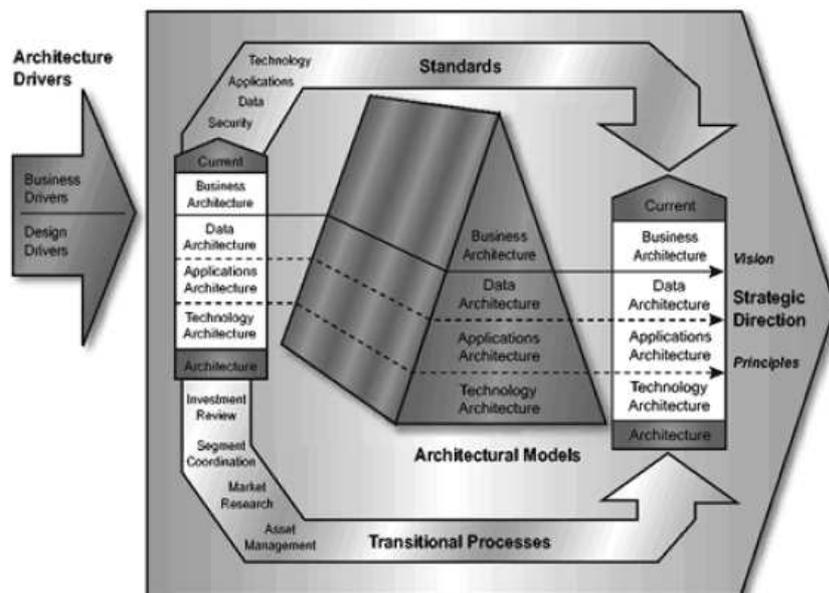


Figura 32 - Composição da Arquitetura FEAF

A Tabela 5 mostra a Matriz da Arquitetura FEAF, cujos os nomes dos produtos da arquitetura corporativa a serem desenvolvidas para cada célula. A versão 1.1 do FEAF não prescreve o conteúdo ou abordagem para o desenvolvimento desses produtos de trabalho.

| | Arquitetura Dados | Arquitetura Aplicação | Arquitetura Tecnológica |
|-------------------------------|-------------------------------|------------------------------|--|
| Perspectiva do Planejamento | Relação de Objetos de Negócio | Relação Processos de Negócio | Relação localização do Negócio |
| Perspectiva do Dono | Modelo Semântico | Modelo de Processo Negócio | Sistema Lógico de Negócio |
| Perspectiva do Projetista | Modelo Lógico Dados | Arquitetura de Aplicação | Arquitetura de desenvolvimento de Sistema Geográfico |
| Perspectiva do Construtor | Modelo Físico Dados | Projeto de Sistemas | Arquitetura Tecnológica |
| Perspectiva do Subcontratante | Diretório Dados | Programas | Arquitetura Rede |

Tabela 8 - Matriz da Arquitetura FEAF

A-IV.6 ISO RM-ODP

O Modelo de Referência ISO para processamento aberto distribuído, vulgarmente designado por RM-ODP, fornece uma estrutura para apoiar o desenvolvimento de padrões que apoiará o processamento distribuído em ambientes heterogêneos. Baseia-se, tanto quanto possível, na utilização de técnicas de descrição formal para a especificação da arquitetura.

RM-ODP usa uma abordagem de modelagem de objetos para descrever sistemas distribuídos. Duas abordagens de estruturação são usadas para simplificar os problemas de modelagem em sistemas grandes e complexos: cinco visões fornecem maneiras diferentes de descrever o sistema, e oito transparências identificam problemas específicos, únicos em sistemas distribuídos, que os padrões de sistemas distribuídos devem resolver. Cada ponto de vista está associado com uma linguagem que pode ser utilizada para descrever os sistemas daquele ponto de vista.

Os cinco pontos de vista descritos pela RM-ODP são:

1. O *ponto de vista da empresa*, que analisa o sistema e seu ambiente no contexto das exigências do negócio no sistema, o seu objetivo, âmbito e políticas. Trata-se de aspectos da organização, tais como sua estrutura organizacional, que afetam o sistema.
2. O *ponto de vista da informação*, que foca na informação tratada no sistema. Como as informações estão estruturadas, como elas se modificam, os fluxos de informação, e as divisões lógicas entre as funções independentes dentro do sistema são todas tratadas neste ponto de vista.
3. O *ponto de vista computacional*, que incide foca na decomposição funcional do sistema em objetos que interagem com interfaces.
4. O *ponto de vista da engenharia*, que se concentra em como interações distribuídas entre os objetos do sistema é suportado.
5. O *ponto de vista da tecnologia*, que se concentra no hardware e componentes de software que compõem o sistema.

A-IV.7 SPIRIT

A base da plataforma do SPIRIT (*Service Providers' Integrated Requirements for Information Technology*) é uma especificação que foi desenvolvida no âmbito do Fórum de Gestão de Rede, hoje conhecida como TMF (*TeleManagement Forum*). A Edição 3.0 foi publicada pelo *The Open Group*. SPIRIT é um esforço conjunto entre os prestadores de serviços de telecomunicações, fornecedores de sistemas computadorizados e fornecedores de software independentes, com o objetivo de produzir uma abordagem comum, conforme um conjunto de especificações para uma plataforma de computação para propósitos genéricos. O objetivo é fornecer um conjunto de especificações para utilização em cada empresa, na compra de componentes de software para plataformas

de computação heterogeneas. As especificações do SPIRIT são essencialmente baseadas nos padrões da indústria e amplamente aceitos.

A-IV.8 TEAF

Em Julho de 2000, o Departamento do Tesouro publicou o TEAF (*Treasury Enterprise Architecture Framework*). O TEAF dispõe de:

1. Orientação aos departamentos do Tesouro relativos ao desenvolvimento e evolução de uma Arquitetura de Sistemas de Informação.
2. Um conceito único, princípios comuns, tecnologias e padrões para os sistemas de informação.
3. Um modelo (*template*) para desenvolvimento da arquitetura empresarial.

O TEAF descreve um *framework* arquitetural que suporta processos de negócio do Tesouro, em termos de produtos de trabalho. Este *framework* orienta o desenvolvimento e remodelagem dos processos de negócios para os diversos escritórios a fim de satisfazer as exigências da recente legislação em um ambiente tecnológico em rápida mutação. O TEAF prevê visões da arquitetura e um conjunto de produtos essenciais e de apoio para retratar essas visões. A Figura 33 ilustra o *framework* TEAF.

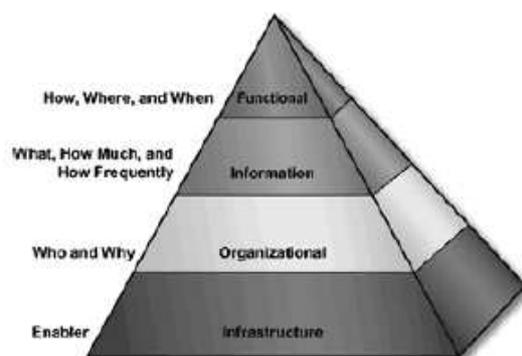


Figura 33 - Framework TEAF

As visões: funcional, informação e organizacional da arquitetura TEAF modelam coletivamente os processos, procedimentos e operações de negócio da organização. Alicerçando a arquitetura no negócio da organização, o TEAF define os procedimentos e processos centrais do negócio da empresa. Através de seus modelos explícitos, a arquitetura TEAF permite a identificação e fundamentação da empresa e preocupações a nível de sistemas e decisões de investimento.

O TEAF separa a arquitetura de informação da empresa em Diretivas da Arquitetura da organização (diretrizes, políticas e roteiro de programa), da Arquitetura Empresarial e Realização da Arquitetura Corporativa (estratégia de transição, previsões técnicas e de

inserção). A Descrição da Arquitetura Empresarial é uma matriz, com colunas sendo vistas sob o enfoque funcional, de informação, organizacional e infra-estrutura e as linhas sob a perspectiva: do planejador, do proprietário, do projetista e do construtor. Muitos dos produtos do TEAF também se encontram no framework C4ISR. No entanto, TEAF introduz um Modelo de Garantia da Veracidade da Informação (*Information Assurance Trust Model*), Administração de Avaliação de Risco da Informação (*Information Assurance Risk Assessment*), e o Roteiro de Arquitetura Empresarial (*Enterprise Architecture Roadmap*). Este roteiro resume o escopo da arquitetura da empresarial, os direcionadores, plano de governança, abordagem e metodologia, e um plano de programa.