

**UNIVERSIDADE ESTÁCIO DE SÁ
FACULDADE INTEGRADA DO CEARÁ**



**CURSO DE ESPECIALIZAÇÃO LATO SENSU EM
GOVERNANÇA DE TI**

ANTONIO ALEX DE SOUZA

**A IMPORTÂNCIA DA QUALIDADE DOS DADOS NAS EMPRESAS COM
ÊNFASE NO APOIO À TOMADA DE DECISÃO:
UM ESTUDO EXPLORATÓRIO**

FORTALEZA (CE)

2017

**A IMPORTÂNCIA DA QUALIDADE DOS DADOS NAS EMPRESAS COM
ÊNFASE NO APOIO À TOMADA DE DECISÃO:
UM ESTUDO EXPLORATÓRIO**

Artigo científico apresentado ao curso de Pós-graduação em Governança de TI pela Faculdade Integrada do Ceará, como requisito para a obtenção do grau de Especialista.

Orientador: Otávio Fernandes Frota.

FORTALEZA (CE)

2017

A IMPORTÂNCIA DA QUALIDADE DOS DADOS NAS EMPRESAS COM ÊNFASE NO APOIO À TOMADA DE DECISÃO: UM ESTUDO EXPLORATÓRIO

Antonio Alex de Souza¹

Otávio Fernandes Frota²

RESUMO

Compreender como os profissionais lidam com dados e informações das empresas e ainda como visualizam a qualidade destes dados é essencial para trilhar estratégias de gerência desta qualidade. Por meio de um questionário online buscou-se observar como profissionais de TI e de áreas relacionadas a dados estão tratando com a questão da qualidade dos dados em suas empresas, principalmente dados que são base para tomada de decisão (consumidos por softwares de business intelligence) e se existe um apoio a esses profissionais por parte da alta gestão das empresas em atividades relacionadas à qualidade de dados. Dentre os resultados obtidos, vimos que os profissionais tem a consciência e sabem da importância da qualidade dos dados e o que ela pode gerar para a empresa, mas não tem o devido apoio por parte da alta gestão. A pesquisa também evidenciou a necessidade de projetos relacionados à qualidade de dados dentro de um programa de Governança de Dados.

Palavras-Chave: dados. Tomada de decisão. Qualidade de dados.

ABSTRACT

Understanding how professionals deal with business data and information, and how they visualize the quality of these data, is essential to driving quality management strategies. Through an online questionnaire, we sought to observe how IT professionals and data-related areas are dealing with the issue of data quality in their companies, mainly data that are the basis for decision-making (consumed by business intelligence software) and if there is support for these professionals by the top management of companies in activities related to data quality. Among the results obtained, we saw that the professionals are aware and know the importance of data quality and what it can generate for the company, but it does not have the proper support from top management. The research also highlighted the need for projects related to data quality within a Data Governance program.

Keywords: Data. Decision making. Quality of data.

¹ Aluno do curso de MBA em Governança de TI do Centro Universitário Estácio do Ceará.

² Professor mestre orientador do artigo.

1 INTRODUÇÃO

Dispor da informação certa, em tempo hábil, 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, ver o melhor caminho, adequar-se às variações da economia e normas regulatórias e ainda se antecipar às necessidades de seus clientes. Uma forma de prover um ambiente para apoiar tais decisões é através de projetos de *Business Intelligence* (BI) [Friedman et al. 2009]. Nem todos os projetos são entregues com êxito ou atendem a todos os requisitos acordados e em projetos de BI não é diferente, um dos principais motivos do insucesso de muitos destes projetos está relacionado à qualidade dos dados.

Dados sem qualidade resultam em relatórios e gráficos imprecisos, análises incorretas e incompletas, duplicidade de informações, dependência excessiva do conhecimento tácito dos tomadores de decisão, entre outros. A qualidade dos dados é crucial para que os tomadores de decisões (gerentes, diretores) tenham como extrair informações úteis e estratégicas para a empresa, entretanto, são poucas as empresas que podem confiar totalmente em seus dados.

Para analisar como uma organização adota práticas de gestão de dados, há uma série de perguntas a serem feitas e, segundo Campos (2005), 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. Também é considerado indicativo o comprometimento com a qualidade dos dados por parte da alta gestão e a existência de uma metodologia para o assunto.

A partir desse ponto de vista, despertou no pesquisador o interesse em compreender como os profissionais de Tecnologia da Informação (TI) e de áreas relacionadas estão lidando com a qualidade dos dados e este se tornou o objetivo principal do trabalho. Os objetivos específicos visam: a) apresentar uma fundamentação teórica dos pontos abordados; b) pesquisar e tabular dados sobre como está sendo tratada a qualidade dos dados nas empresas brasileiras; c) apresentar os resultados obtidos com a pesquisa.

Seguindo-se a esta introdução, tem-se a metodologia da pesquisa em sua classificação e tipologia. Em seguida, apresenta-se a governança de dados em seus conceitos e áreas de conhecimento, assim como detalhou-se a área de qualidade de dados. A sessão seguinte apresenta um estudo de caso, onde foi aplicado um questionário online com o intuito

de observar como a qualidade dos dados está sendo tratada nas empresas e, logo em seguida, analisou-se os resultados obtidos com a aplicação do questionário.

O artigo se encerra com a apresentação das principais conclusões, bem como outras perspectivas para a realização de futuros trabalhos.

2 METODOLOGIA DE PESQUISA

A metodologia utilizada em uma pesquisa traduz a forma como foi desenvolvida. No caso específico deste estudo, quanto aos meios ou aos procedimentos técnicos, foi bibliográfica, cujos principais assuntos abordados foram governança de dados e como as empresas brasileiras estão tratando a qualidade dos seus dados. Exemplificando esse tipo de pesquisa, tem-se que:

[...] não deve ser confundida, como acontece frequentemente, com a pesquisa de documentos. O levantamento bibliográfico é mais amplo do que a pesquisa documental, embora possa ser realizado simultaneamente com a pesquisa de campo e de laboratório. A pesquisa bibliográfica tem por finalidade conhecer as diferentes formas de contribuição científica que se realizaram sobre determinado assunto ou fenômeno. Normalmente, o levantamento bibliográfico é realizado em bibliotecas públicas, faculdades, universidades e, especialmente, naqueles acervos que fazem parte do catálogo coletivo e das bibliotecas virtuais. (OLIVEIRA, 2010, p. 119).

Bibliográfica, pois, segundo Vergara (2010, p. 43), foi realizada a partir de material teórico já existente em livros, revistas, jornais, artigos e sites da internet, utilizando dados secundários, ou seja, material já publicado anteriormente.

Do ponto de vista de seus objetivos, a pesquisa pode ser considerada descritiva e exploratória. Descritiva, por descrever os principais aspectos inerentes à governança e qualidade de dados. Também exploratória, pois através de um questionário online colheram-se dados reais referentes a como as empresas estão tratando a qualidade dos seus dados. As pesquisas exploratórias, segundo Gil (2002), são desenvolvidas com o objetivo de proporcionar visão geral, de tipo aproximativo, acerca de determinado fato. Vergara (2009) afirma que uma investigação exploratória é realizada em área na qual há pouco conhecimento acumulado e sistematizado.

Andrade (2008) 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 vista 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.

3 FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica está dividida em dois tópicos, sendo eles: governança de dados – conceitos e suas áreas; qualidade de dados.

3.1 Governança de dados

A capacidade que uma organização possui em proteger seus dados, revesti-los de qualidade e produzir informações confiáveis, precisas, acessíveis e disponíveis no momento correto é um dos principais fatores determinantes para o valor das organizações modernas. (FERNANDES; ABREU, 2012).

A Governança de Dados (GD) tem como responsabilidade o controle e o gerenciamento dos dados da organização, sendo a responsável por todo controle dos dados e a transformação dos mesmos em informações. A GD é um tema novo que vem ganhando destaque e importância nas organizações, principalmente em organizações onde existe uma grande massa de dados e estes são de extrema importância para a tomada de decisões.

Segundo o *Data Governance Institute* (*Data Governance Institute*, n.d.), governança de dados é um procedimento de tomada de decisões e responsabilidades para com os processos relacionados aos dados, baseando-se em políticas, normas e restrições. O foco de atuação da GD pode variar de organização para organização, começando em privacidade de dados e indo até a qualidade dos mesmos, passando por diversas áreas do ciclo do dado. Para possuir uma GD estruturada e eficiente, é preciso que as organizações definam suas necessidades de gestão de dados, bem como os objetivos a serem atingidos, e a partir deste ponto delimitem o escopo de atuação da GD. (FERNANDES; ABREU, 2012).

Conforme o *Data Governance Institute*, a Gestão de Dados possui seis áreas-focos:

- 1) Políticas, Normas e Estratégias;
- 2) Qualidade dos dados;
- 3) Privacidade/*Compliance/Security*;
- 4) Arquitetura/Integração;

5) *Data Warehouse* (DW) e *Business Intelligence* (BI);

6) Alinhamento entre governança de dados e as estratégias de TI e negócio;

Toda organização que englobar estes seis grandes temas em sua Gestão de Dados, de maneira estruturada e eficiente, alinhados aos objetivos da organização, provavelmente possuirá um maior sucesso em sua gestão de dados, bem como proporcionará dados com maior relevância, os quais serão utilizados de maneira correta para as tomadas de decisões estratégicas da organização.

Segundo Mosley et al. (2009) governança de dados é o exercício da autoridade, do controle e da tomada de decisão sobre os ativos de dados, ou seja, a GD é responsável pelo gerenciamento deles e ainda pelo planejamento, supervisão, controle e seu uso. A GD divide-se em dois grandes grupos de atividades: (1) planejamento do gerenciamento de dados e (2) controle do gerenciamento de dados.

Por meio destas atividades de GD, é possível se estruturar o gerenciamento de dados de maneira eficiente. Entretanto, não basta seguir as atividades e os objetivos relatados, é necessária a utilização de *frameworks* que auxiliem no desenvolvimento e implementação de políticas de GD, para que esta seja bem sucedida e gere valor, tanto para os dados da organização quanto para as decisões estratégicas futuras que serão tomadas.

Das áreas focos expostas acima, irar-se detalhar, logo em seguida, a área de qualidade de dados, que é o assunto principal do trabalho.

3.1 Qualidade de dados

Conforme Barbieri (2013), o objetivo é planejar, implementar e controlar atividades que apliquem técnicas de gerência de qualidade de dados para medir, avaliar, melhorar e garantir a adequação dos dados ao seu uso pretendido. A estrutura da atividade desta função é:

- Desenvolver e promover aspectos de conscientização sobre Qualidade de Dados – aqui a grande questão é vender a importância da qualidade de dados nas empresas. É preciso difundir a importância dos conceitos, seja por mecanismos diretos ou indiretos. Mostrar também que o problema não é (somente) do domínio da TI, mas principalmente um problema da esfera de negócios. A qualidade de dados deve ser um dos elementos fundamentais do arco da Governança de Dados da empresa, que define política, padrões, procedimentos, papéis, programas e projetos dentre outros itens do seu escopo, visando tratar e preservar ao ativo

“dado”. A realização de um trabalho inicial de *profiling* dos dados mais importantes da empresa, mostrando os resultados preocupantes com relação à qualidade dos dados é algo a ser fortemente pensado, pois serve como “*start-up*” para todo o processo de convencimento material sobre os problemas de dados.

- Definir requisitos de qualidade de dados – os requisitos de qualidade de dados são definidos diretamente em função das necessidades da empresa. Há que se pensar nos processos críticos da empresa, suas regras de negócios, seus dados consumidos e produzidos e o impacto da qualidade dos dados na sua execução, tanto como input quanto output. Esse é o início de tudo. Os requisitos de qualidade de dados passam por vários domínios, que podem variar de acordo com os autores. O mostrado pelo DMBOK é:

- Precisão (*accuracy*) ou como as “coisas”/entidades da vida real estão corretamente representadas;
- Completude (*completeness*): o quão completos estão os dados (todos os atributos? Faltam alguns? Todos os essenciais? Alguns acessórios?) exigidos na execução daquele processo de negócio;
- Consistência (*consistency*): se refere à integridade cruzada entre duas ou mais fontes que armazenam o mesmo dado. Há coerência entre esses dados que habitam fontes diferentes? A coerência existe no mesmo contexto ou em contextos diferentes?
- Atualidade (*currency*): o quanto os dados estão atualizados e representam o estado corrente e mais atual;
- Precisão numérica (*accuracy*): representação de valores no grau de precisão necessária, como casas decimais para dados numéricos;
- Disponibilidade (*availability*): o dado é disponibilizado no momento de sua necessidade?
- Unicidade (*uniqueness*): o fato de haver representação única de certa entidade, sem ambiguidade ou sentidos diferentes.

- Estabelecer processos de “*profiling*”, análise e avaliação de Qualidade de Dados – refere-se à necessária fotografia inicial do estado dos dados de certa(s) área(s) de assunto ou domínio(s) da empresa. Deve ser uma das primeiras ações para se estabelecer as “*baselines*” dos processos de melhoria de qualidade de dados da empresa. Permite criar as primeiras métricas e a definição dos objetivos a serem alcançados em função delas. É como se fosse a análise laboratorial solicitada por um médico para melhor diagnosticar o estado do paciente e iniciar o seu tratamento.

- Definir métricas para Qualidade de Dados – as métricas deverão ser definidas para a avaliação do estado atual e da evolução dos tratamentos de qualidade dos dados. As métricas, como todas as medidas definidas em processos de qualidade, deverão: ser atreladas a objetivos bem definidos; responder a questões associadas a esses objetivos e ser medidas definidas com clareza, que apontem elementos quantificáveis associáveis a objetivos de negócios, com formulações claras (como medir), valores definidos para análise (como analisar), com faixas aceitáveis e não aceitáveis (como interpretar), plano de ações no caso de discrepâncias, frequência de medição (quando medir), entre outros.

- Definir regras de negócios para Qualidade de Dados – implica na análise das regras de negócios fundamentais dos processos e na descoberta dos dados que podem implicar em eventuais quebras de conformidade delas. Esses dados deverão ser observados na sua qualidade justamente para garantir a conformidade da regra com o processo. Por exemplo, a regra de negócios define que nenhum colaborador com oito horas diárias de turno de trabalho poderá ganhar menos que o salário mínimo. Um campo de um arquivo enviado ao INSS contendo um valor abaixo desse estabelecido implica numa quebra de conformidade do processo (admissão, por exemplo), com as regras de negócios definidas.

- Testar e validar os requisitos de Qualidade de Dados – nesse ponto, o DMBOK sugere que haja processo de verificação inicial (*data profiling*, por exemplo) e verificação constante e recorrente, a fim de que os dados sejam sempre avaliados nos seus domínios de qualidade.

- Definir e avaliar níveis de serviços de Qualidade de Dados – nesse item, é sugerida a definição de níveis de serviços de qualidade de dados, o que deverá ser garantido por medições e verificações constantes. Os níveis de SLA são o compromisso firmado sobre qualidade da área gestora dos dados com os seus usuários. Os itens subsequentes, relativos a medir e monitorar continuamente, gerenciar as pendências e corrigir os defeitos são consequências diretas desse compromisso de nível de serviços.

- Medir e monitorar continuamente a Qualidade de Dados – está relacionado aos níveis de granularidade de medição de dados: valor do elemento de dados; instância de dados do registro; pacote de dados.

- Gerenciar as pendências de Qualidade de Dados – passos para catalogar problemas com qualidade de dados: padronizar atividades e problemas; designar um processo para problemas com dados; gerenciar problemas com processos de escala; gerenciar problemas com o ritmo de trabalho.

- Corrigir os defeitos de Qualidade de Dados – atividades operacionais desta etapa: detectar e eliminar a raiz do problema; isolar os itens de dados que estão incorretos. Técnicas de limpeza: correção automatizada; correção manual direcionada (automatizada com revisão manual); correção manual.

- Projetar e implementar procedimentos operacionais da Gerência de Qualidade de Dados – atividades relacionadas a esta etapa: inspeção e monitoramento; diagnóstico e avaliação de alternativas de reparos; resolução do problema; relatórios.

- Monitorar os procedimentos operacionais e a performance da Gerência de Qualidade de Dados – principais passos dessa etapa: atribuir responsabilidade; evitar zonas cinzentas onde ninguém atua, métricas são indicadores ao ritmo de trabalho; métricas são indicadores para diagnóstico, planejamento e solução de problemas.

4 PESQUISA DE CAMPO

Esta pesquisa de campo tem o objetivo de compreender como os profissionais de Tecnologia da Informação (TI) e áreas relacionadas, de empresas brasileiras, estão lidando com a questão da qualidade de dados, principalmente dados que são base para a tomada de decisão, como também avaliar a necessidade de apoio da alta gestão em atividades relacionadas à gerência da qualidade.

Tal compreensão permite traçarem-se as melhores estratégias para prover a qualidade de dados, seja através da definição de critérios de qualidade, metodologias, sistemáticas ou boas práticas a serem adotadas nas empresas ou mesmo a própria necessidade de estratégias de conscientização e uso da governança de dados, como observado nos resultados do estudo aqui realizado.

A coleta dos dados foi realizada através de um questionário online (apêndice I) disponibilizado na internet. A escolha desse instrumento de pesquisa é justificada pela conveniência ao entrevistado em responder no momento e local que julgar mais apropriado, não tendo influência de opiniões e do aspecto pessoal do pesquisador. O questionário é de múltiplas escolhas, com uma série de respostas possíveis.

O questionário foi divulgado através de aplicativo de mensagens instantâneas a grupos voltados a profissionais de TI, mas especificamente profissionais de dados e de negócio (Administradores de Banco de Dados, Analista de Dados, Analistas de *Business Intelligence*, Cientistas de Dados, Gerentes de TI, Coordenadores de TI, Consultores de TI e

Analistas de Negócios) de empresas dos diversos ramos de negócio, sediadas no Brasil. Em um segundo momento, foi divulgado no blog do pesquisador e posteriormente publicado em uma rede social de uso estritamente profissional.

Houve colaboração de alguns profissionais no sentido de indicarem outros profissionais envolvidos em projetos de apoio à decisão e que poderiam estar participando mais efetivamente da pesquisa. Esta pesquisa exploratória foi respondida por 164 profissionais de TI e de áreas relacionadas, distribuídos em 15 estados, contemplando todas as regiões do Brasil. Este questionário ficou disponível por 16 dias e o período da coleta dos dados ocorreu entre os dias 27 de dezembro de 2016 e 11 de janeiro de 2017.

No que se refere ao perfil dos 164 profissionais que responderam à pesquisa, podemos dizer que aproximadamente 82% dos respondentes são dos estados do Ceará e de São Paulo, com isso, a pesquisa tendeu a mostrar mais a realidade destes dois estados. Observou-se também que 93% dos respondentes são da área de TI e 7% de áreas relacionadas; dentre os profissionais de TI, 34% possuem tempo de experiência de até dois anos, 24% de três a quatro anos e 42% entre cinco e vinte anos no cargo. Em relação a tempo de experiência percebeu-se que a maioria possui mais de três anos de experiência em suas respectivas funções e, com isso, tendem a possuir um bom domínio da área de atuação e propriedade em suas respostas.

Já em relação aos cargos, estes foram divididos por subáreas de conhecimento (tabela 1), para que houvesse uma ideia melhor do público respondente da pesquisa; são elas: Profissionais de Dados (34%), Gestão ou Negócio (42%), Desenvolvimento (6%), Técnico (11%) e Relacionado (7%). Podemos observar que mais de 76% dos respondentes são profissionais que trabalham diretamente com dados e/ou estão ligados diretamente a áreas estratégicas das empresas.

Cargo	Profissionais de Dados	Gestão ou Negócio	Desenvolvimento	Técnico	Relacionado
DBA	X				
Analista de Banco de Dados	X				
Analista de Dados	X				
Cientista de Dados	X				
Analista de Negócio		X			
Analistas de BI		X			
Analista de Projetos		X			
Analista de Sistemas		X			
Analista de TI		X			
Coordenador de TI		X			
Consultor de TI		X			

Gerentes de TI		X			
Diretor de TI		X			
Supervisor de TI		X			
Desenvolvedores			X		
Analista de Suporte				X	
Técnico de Infra				X	
Analista de Infra				X	
Analista de Segurança				X	
Consultor					X
Estatístico					X
Gerente(s)					X
Professor					X
Analista(s)					X
Controle de Qualidade					X
Secretário					X
Supervisor					X

Tabela 1 Subáreas de conhecimento

Fonte: autoria própria (2017).

5 RESULTADOS OBTIDOS

Quanto a patrocínio de projetos relacionados à melhoria contínua na qualidade dos dados por parte da alta gestão (diretoria e gerência) da empresa, observa-se que infelizmente apenas 35% identificaram este envolvimento (sempre/com frequência) e os demais afirmaram que tal fato raramente acontece (44%), nunca acontece (13%) ou não sabem informar (8%). Desta forma, podemos observar uma falta de conscientização da importância da qualidade dos dados.



Figura 1 Identificam profissionais responsáveis por ações de melhoria na qualidade dos dados transacionais

Fonte: autoria própria (2017).

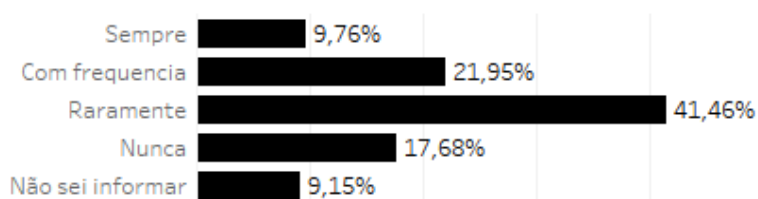


Figura 2 Identificam profissionais pertencentes às áreas de negócio da organização envolvidas com ações de melhoria na qualidade dos dados

Fonte: autoria própria (2017).

No entanto, observa-se que um pouco mais de 57% dos respondentes conseguiram identificar na empresa profissionais responsáveis por ações de melhoria na qualidade dos dados transacionais (figura 1), mas, aproximadamente apenas 32% dos respondentes informaram que nas empresas onde trabalham profissionais das áreas de negócio da organização estão sempre, ou com frequência, envolvidos em ações que levem à melhoria na qualidade dos dados (figura 2).

Esses números relativamente baixos em relação à identificação dos responsáveis por ações de melhoria na qualidade de dados, assim como o baixo envolvimento de profissionais das diversas áreas de negócio, nos reforçam a ideia de que o assunto ainda não tem a devida importância dentro das empresas e, por este motivo, projetos e ações voltadas à melhoria na qualidade de dados são praticamente inexistentes.

Grande parte dos que responderam à pesquisa, ou seja, 77%, reconhecem que o uso dos dados com qualidade permitirá que a organização seja mais eficiente. Já 65% acham que ela se torna mais ágil, 64% mais competitiva, 63% ainda mais organizada (contra 4% que acreditam que ela se tornará mais burocrática), 48% acreditam ainda que a organização se tornará mais lucrativa, sugerindo assim um bom nível de conscientização do profissional acerca dos ganhos efetivos que podem ser obtidos pela empresa com a utilização dos dados com qualidade.

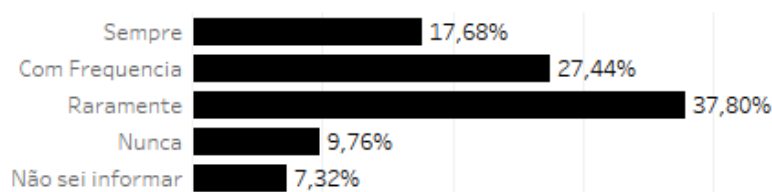


Figura 3 Documentação dos modelos e fontes de dados

Fonte: autoria própria (2017).

Vemos (figura 3) que um pouco mais de 45% dos pesquisados afirmaram que, sempre ou com frequência, a documentação dos modelos e fontes de dados são mantidas pela equipe de TI, o que indica certo descaso com a documentação dos dados e, portanto, uma possível ingerência relacionada à qualidade dos dados.

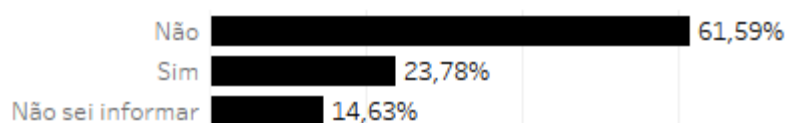


Figura 4 Processos que medem a qualidade dos dados

Fonte: autoria própria (2017).

Na figura 4, pôde-se observar que apenas 24% dos respondentes afirmaram existir um processo que avalia a qualidade de dados existentes na organização, isso indica, mais uma vez, à questão de que é preciso uma conscientização da importância dos dados e de manter processos relacionados à qualidade dos mesmos.

Na literatura pode-se encontrar várias metodologias para a definição, medição, análise e melhoria da qualidade dos dados; a adoção de uma destas metodologias tende a possibilitar uma maior gerência da qualidade dos dados na empresa. Na pergunta específica para se verificar a utilização de alguma metodologia, foram destacadas quatro (que em geral são as mais utilizadas): ISO 9001 [ISO 2008], GQM [Basili et al. 1994] , TDQM [Wang 1998] e DMBOK [Barbieri 2013]. Porém, um pouco mais de 60% afirmam que não utilizam nenhuma metodologia que organiza conceitos e procedimentos para a definição, medição, análise e melhoria da qualidade dos dados (13% utilizam ISO 9001 e menos de 4% utilizam TDQM, GQM ou DMBOK).

Levando-se em consideração as várias propostas de metodologia, como também os inúmeros critérios de qualidade de dados, existia a expectativa da adoção de alguma metodologia e até mesmo a dificuldade em escolher uma metodologia e critérios mais adequados para a necessidade da empresa, mas os resultados indicaram um desconhecimento da importância do uso de uma metodologia e definição adequada de critérios; isto mostra que este problema vai além do âmbito técnico.

Em seguida, foram realizadas perguntas relacionadas aos tipos de tratamento de dados efetuados nas fases do processo de limpeza (antes de ir para um software de BI, por exemplo) e o quanto dos problemas encontrados são efetivamente tratados (figura 5). Aproximadamente 33% dos respondentes indicaram que conflitos de domínios ou necessidade de conversão nunca, ou raramente, são tratados nesse processo (e outros 24% não sabem informar). Isso mostra claramente problemas relacionados à qualidade dos dados, pois uma vez encontrados problemas neste sentido, deveriam ser devidamente documentados e encaminhados aos responsáveis, como, por exemplo: para a equipe ou empresa responsável pelo desenvolvimento do ERP, isto tornaria os dados mais confiáveis para relatórios e gráficos gerenciais voltados à tomada de decisão.

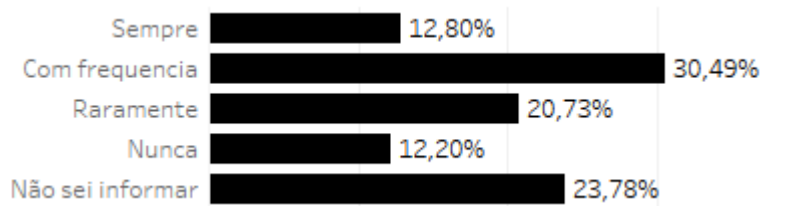


Figura 5 Conflito de domínios ou necessidade de conversão de atributos no processo de ETL

Fonte: autoria própria (2017).

Já em relação a fontes de dados com regras de negócio inconsistentes e violações de regras de integridade nas fontes de dados, pôde-se observar nas figuras 6 e 7 que apenas 40%, em média, das empresas conseguem identificar estas inconsistências e violações durante o processo. Esse processo deve ser muito bem documentado e encaminhado para análise/solução pelos devidos responsáveis, por exemplo: setor de desenvolvimento, empresa responsável pelo ERP, entre outras. O que preocupa é que os números são ainda muito baixos.

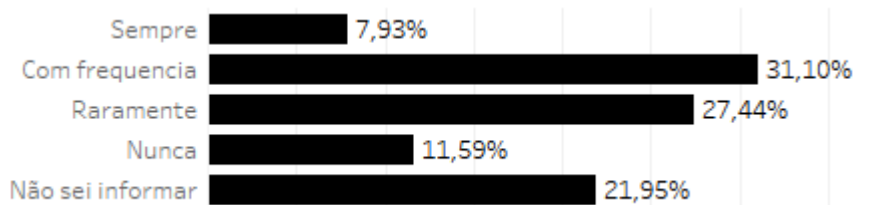


Figura 6 Inconsistência de regras de negócio no processo de ETL

Fonte: autoria própria (2017).

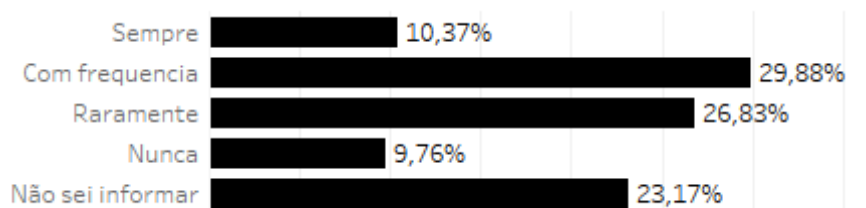


Figura 7 Violação de integridade em fonte de dados

Fonte: autoria própria (2017).

Aproximadamente 46% dos pesquisados reconheceram que o processo de tratamento e limpeza dos dados, expostos nas figuras 5, 6 e 7 acima, devem gerar correção no sistema de origem, para que posteriormente possam ser carregados para um software de *Business Intelligence*, como podemos ver logo mais abaixo na figura 8. Mas, na realidade, devido a diversos fatores, como, por exemplo: sistemas não documentados e complexos, sistemas sem suporte e manutenção, descontinuados pelo fabricante, falta de profissional especializado etc. estes tipos de problema são tratados no momento da carga pelo processo de ETL, gerando um grande retrabalho, pois em cada carga realizada, o dado deverá ser tratado.

Isso é o relato de vários colegas consultores da área de dados, que se deparam com situações como essas, corriqueiramente.

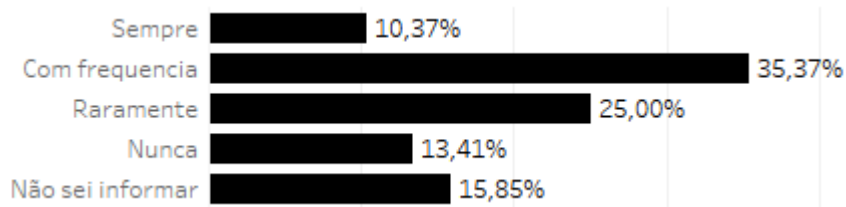


Figura 8 Avaliação do tratamento de dados que geram correções na origem

Fonte: autoria própria (2017).

Quase 37% dos respondentes acreditam que o processo de tratamento e limpeza dos dados garante cerca de 80 a 100% da qualidade de dados, enquanto 34% acredita que esta garantia ficaria em torno de 60 a 80% e, com isso, garante-se dados mais confiáveis para apoio à tomada de decisões (figura 9). Com esse indicador, podemos constatar que os profissionais percebem a importância e o que a qualidade de dados pode gerar de benefício à empresa.

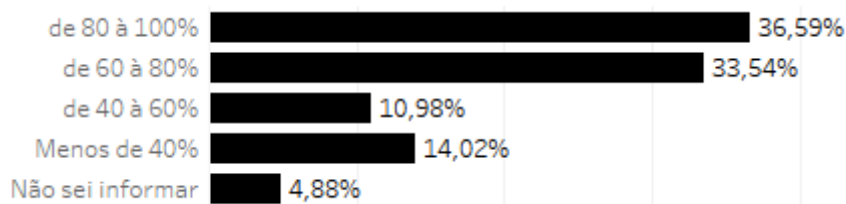


Figura 9 Avaliação da garantia da qualidade de dados executada pelo processo de tratamento e limpeza dos dados

Fonte: autoria própria (2017).

Apesar de termos no questionário da pesquisa a opção de 100% de garantia na qualidade dos dados, todos nós temos consciência de que nunca será possível garantirmos 100% em função do grande volume de dados com que as empresas lidam diariamente, ainda mais com o advento do BIG Data e o dinamismo dos negócios, diante do mercado cada vez mais competitivo.

6 CONCLUSÕES

Este artigo teve como objetivo analisar através de uma pesquisa exploratória como as empresas estão tratando o seu maior patrimônio, seus dados. Com o resultado obtido, viu-se que a qualidade dos dados nas empresas brasileiras precisa urgentemente de atenção,

devido à maioria dos profissionais das diversas áreas não terem a noção da importância de processos relacionados à qualidade de dados e do que podem trazer de retorno para a empresa, principalmente no que se refere à tomada de decisão.

Nota-se que o assunto qualidade de dados ainda é tratado 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 todas as inconsistências e problemas de qualidade de dados está no processo de extração, transformação e carga para um software de BI, por parte de profissionais de TI. Em função destes equívocos é que o tratamento dos dados e informações deve ser reconhecido e tratado como qualquer outro recurso tangível existente na empresa.

Essencialmente, a pesquisa evidenciou que boa parte dos profissionais tem consciência, ou conhecimento da importância de um processo de qualidade de dados e o que isto gera de benefícios para as empresas, principalmente quando relacionada a dados para a tomada de decisão. Mas não existe o apoio necessário da alta gestão em projetos e ações neste sentido, como, por exemplo: projeto de documentação das fontes de dados; uso de metodologias e processos que medem a qualidade dos dados, dentre outros, não conseguem colocar as boas práticas em ação. Como também se identifica claramente a necessidade da inserção de projetos relacionados à qualidade de dados dentro de um programa de Governança de Dados e a conscientização, principalmente da alta gestão, em relação a esta problemática.

Algumas sugestões de trabalhos futuros: (a) uma estudo detalhado da qualidade de dados nas empresas no que se diz respeito ao apoio à decisão em um universo maior de empresas e de profissionais; (b) uma análise sobre a relação de causa e efeito entre a qualidade dos dados utilizados para a tomada de decisões e a qualidade das decisões que são tomadas no ambiente empresarial e (c) a definição e propagação de um programa de Governança de Dados, em nível nacional, que possibilite contribuir com as empresas na manipulação de dados de forma mais efetiva, objetivando prover definições consistentes, prescrever uma administração de dados na empresa, medir e monitorar a qualidade dos dados transacionais e analíticos utilizados nas suas diversas áreas.

7 REFERÊNCIAS

- ANDRADE, Maria Margarida. **Como preparar trabalhos para cursos de pós-graduação**. 7ª ed. São Paulo: Atlas, 2008.
- BARBIERI, Carlos. **Uma visão sintética e comentada do Data Management Body of Knowledge (DMBOK)**. FUMSOFT, Belo Horizonte, 2013.
- BASILI, Vitor R., CALDIERA, Gianluigi, ROMBACH, H. Dieter. **Goal question metric paradigm**. Encyclopedia of Software Engineering, pp. 528-532, John Wiley & Sons, 1994.
- CAMPOS, A. L. de S. **Indicadores de qualidade de dados em data warehouse**. Monografia de conclusão de curso. UFJF, Juiz de Fora, 48f, 2005.
- CHAUDHURI, S., DAYAL, U. **Requirements capture and analysis: a survey of current practice**. Requirements engineering, v. 1, n. 2, pp. 75-87. Springer-Verlag, 1996.
- Data Governance Institute. (n.d.). Data Governance. Retrieved September 12, 2011, from http://www.datagovernance.com/adg_data_governance_definition/
- FERNANDES, Aguinaldo Aragon; ABREU, Vladimir Ferraz. **Implantando a governança de TI, da estratégia à gestão de processos e serviços**. 3ª ed. São Paulo: Brasport Livros e Multimídia Ltda., 2012, 615p.
- FRIEDMAN, Ted, DECKER, John E. Van. **Gartner FEI technology study reveals finance managers' perspectives on data quality**. GARTNER GROUP REPORT, ID Number: G00169706, 2009.
- GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 4ª edição, 2002.
- International Organization for Standardization. **ISO 9001:2008: Quality Management Systems – Requirements**. Beuth, Berlin, 2008.
- MOSLEY, Mark et al.(2009). **The DAMA Guide to the data management body of knowlegde**. 1stedition, USA, Technics Publications, LLC, 406 p.
- OLIVEIRA, Silvio Luiz de. **Tratado de metodologia científica**. São Paulo: Pioneira, 2010.
- VERGARA, Sylvia Constant. **Projetos e relatórios de pesquisa em administração**. São Paulo: Atlas, 2010.
- WANG, Richard. **A product perspective on total data quality management**. Communications of the ACM, v. 41, n. 2, pp. 58-65, 1998.

APÊNDICE I

Modelo da pesquisa online disponibilizada para levantamento de dados do estudo.

Endereço de e-mail:

Nome:

Cidade:

Estado:

Cargo (ou cargo anterior):

Quanto tempo no cargo:

Há quanto tempo trabalha (trabalhou) na empresa?

☐ Menos de 1 Ano ☐ de 1 à 3 anos ☐ de 3 à 6 anos ☐ Mais de 6 anos

Em relação à Qualidade dos dados de sua empresa, considera ou nota que...

1. Um ou mais gerentes ou diretores patrocinam e incentivam projetos relacionados à governança de dados, como por exemplo, qualidade de dados.

☐ Sempre ☐ Com Frequência ☐ Raramente ☐ Nunca ☐ Não sei informar ☐ Outro

2. Um ou mais colaboradores na sua empresa são responsáveis por ações de melhoria na qualidade dos dados transacionais.

☐ Sim ☐ Não ☐ Não sei informar ☐ Outro

3. Utilizam alguma das metodologias abaixo para definição, medição, análise e melhoria da qualidade dos dados.

☐ DMBOK ☐ TDQM (Total Data Quality Management) ☐ GQM (Goal-question Metric) ☐ ISO 9001 ☐ Não utilizamos ☐ Não sei informar ☐ Outro

4. Colaboradores das áreas de negócio da empresa estão envolvidos em ações de melhoria na qualidade dos dados.

☐ Sempre ☐ Com Frequência ☐ Raramente ☐ Nunca ☐ Não sei informar ☐ Outro

5. A documentação dos modelos e fontes de dados são mantidas e atualizadas pela equipe de TI.

☐ Sempre ☐ Com Frequência ☐ Raramente ☐ Nunca ☐ Não sei informar ☐ Outro

6. Existem processos que medem a qualidade dos dados existentes na empresa.

☐ Sim ☐ Não ☐ Não sei informar ☐ Outro

7. Em sua opinião, a utilização de dados com qualidade permitirá que a sua empresa seja ainda mais:

☐ Ágil ☐ Competitiva ☐ Lucrativa ☐ Eficiente ☐ Organizada ☐ Burocrática ☐ Outro

8. Atributos (campos) que possuem conflitos de domínios ou necessidade de conversão são tratados no processo de limpeza, antes de ir para um BI (Business Intelligence), por exemplo.

☐ Sempre ☐ Com Frequência ☐ Raramente ☐ Nunca ☐ Não sei informar ☐ Outro

9. Inconsistências de regras de negócios em fontes de dados são facilmente identificadas no processo de ETL (Extração, Tratamento e Carga para um BI, por exemplo).

☐ Sempre ☐ Com Frequência ☐ Raramente ☐ Nunca ☐ Não sei informar ☐ Outro

10. Violações de integridade em fontes de dados são facilmente identificadas no processo de ETL (Extração, Tratamento e Carga para um BI, por exemplo).

☐ Sempre ☐ Com Frequência ☐ Raramente ☐ Nunca ☐ Não sei informar ☐ Outro

11. O processo de tratamento e limpeza dos dados geram correções nos sistemas de origem dos dados para que posteriormente possa ser carregado por um software de BI, por exemplo.

☐ Sempre ☐ Com Frequência ☐ Raramente ☐ Nunca ☐ Não sei informar ☐ Outro

12. O processo de tratamento e limpeza dos dados garante a qualidade dos dados

☐ de 80 a 100% ☐ de 60 à 80% ☐ de 40 a 60% ☐ Menor que 40% ☐ Outro