

# Qualidade dos Dados nas Organizações sob o Enfoque de Apoio a Decisão: Um estudo exploratório

Isabel Maria F. dos Santos, Alexandre A. Andreatta, Sean W. M. Siqueira

<sup>1</sup>Departamento de Informática Aplicada – Universidade Federal do Estado do Rio de Janeiro (UNIRIO)  
Rio de Janeiro – RJ – Brazil

{isabel.santos, andreatt, sean}@uniriotec.br

**Abstract.** *Understanding how the professionals of the organization are dealing with data and how they perceive data quality is crucial to define strategies for managing this quality. Through an online form, it was possible to observe how IT professionals are dealing with data quality, particularly those involved in decision support projects and if there is a need for supporting those team in activities related to data quality management and making quality criteria explicit in data. From the results, the research also pointed out the need for considering data quality projects in a program for data governance.*

**Resumo.** *Entender como os profissionais da empresa estão lidando com os dados e como enxergam a qualidade dos dados é crucial para se traçar estratégias de gerência desta qualidade. Através de um formulário online buscou-se observar 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. Dentre os resultados obtidos, a pesquisa evidenciou a necessidade da inserção de projetos relacionados à qualidade de dados dentro de um programa de Governança de Dados*

## 1. Introdução

Disponer 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. Uma forma de prover um ambiente para apoiar tais decisões é através de projetos de *Business Intelligence* (BI) [Friedman *et al.* 2009]. Entretanto, ainda é muito alto o índice de tais projetos que não obtêm êxito ou não atendem as expectativas e necessidades da organização.

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.

Deste modo, 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. 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 investigadas [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. 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.

A partir desta perspectiva, surgiu então o interesse em se verificar como os profissionais de Tecnologia da Informação (TI) (mais precisamente os envolvidos em dados para apoio à decisão) estavam lidando com a questão da qualidade dos dados. Entender tal cenário nas organizações permite traçar melhores estratégias para prover a qualidade de dados, seja através da definição de critérios de qualidade, metodologias, sistemáticas ou boas práticas a serem adotados nas organizações ou mesmo a própria necessidade de estratégias de conscientização e governança de dados, como observado nos resultados do estudo realizado.

Portanto, aplicou-se um questionário online com o intuito de avaliar como a qualidade da informação estava sendo tratada nas organizações brasileiras, e também avaliar a necessidade de apoiar equipes de TI na explicitação de critérios de qualidade nos dados (especificamente em *Data Warehouse* - DW e/ou *Data Marts* - DMs, repositórios de dados típicos de ambientes de apoio a decisão).

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 artigo serão discutidos a metodologia, o contexto e a forma de aplicação do questionário desta pesquisa (Seção 2). Também se encontram descritos os resultados obtidos a partir da aplicação deste questionário (Seção 3), bem como o perfil dos respondentes (Seção 2), com o intuito de analisar as 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. Finalmente, as conclusões são apresentadas na Seção 4.

## 2. Questionário, forma de aplicação e perfil dos respondentes

Após estudo pormenorizado das propostas existentes na literatura sobre qualidade de dados, em particular em ambientes de apoio a decisão (envolvendo DW/DMs), partiu-se para um estudo exploratório. 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 DW e/ou DMs.

Como procedimento metodológico adotado foi utilizada a pesquisa exploratória, pois esta vai além da descrição das características, analisando e explicando porque ou como os fatos estão acontecendo. As pesquisas exploratórias, segundo Gil (2002), são desenvolvidas com 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 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.

O contato inicial foi feito via email para 60 profissionais de TI de organizações de diversos ramos de negócio, sediadas no Rio de Janeiro, que fazem parte da rede de contatos profissionais do 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 possuírem nenhum projeto e nem ferramentas de qualidade de dados; e 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.

A coleta de dados foi feita por meio de um questionário online (Apêndice 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. Foi utilizado um questionário online de múltiplas escolhas fechadas com uma série de respostas possíveis (Apêndice 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. 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.

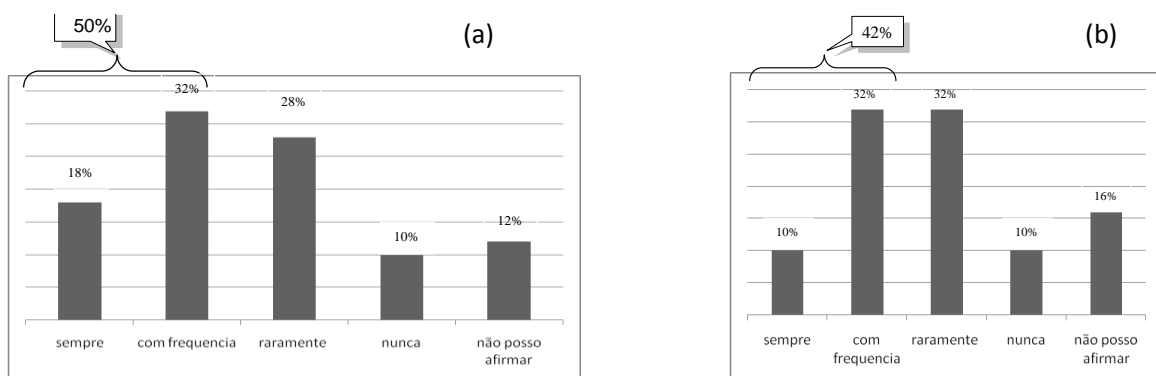
Inicialmente, este questionário ficou disponível por 40 dias, porém este prazo foi estendido por mais 30 dias em função de ter sido disponibilizado em um período de festas Natalinas e férias escolares (muitos profissionais tiram férias em janeiro para viajar com a família). O período de coleta de dados 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. Tal e-mail foi encaminhado após um primeiro contato, explicando o objetivo da pesquisa e solicitando a autorização e colaboração do mesmo.

Com relação ao perfil dos profissionais que responderam à pesquisa, totalizando 43, cerca de 75% possuem experiência acima de dez anos na área de TI (cerca de 10% possuem experiência de cinco anos e o restante com experiência de seis, oito e nove anos). Observou-se também que cerca de 75% dos profissionais são prestadores de serviço. Porém, quando se verifica o tempo em que atuam na área de DW e/ou DM, essa distribuição fica mais uniforme (31% com até dois anos de experiência, 33% de quatro a seis anos e 36% com oito anos ou mais de atuação nesta área).

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 (36% tiveram seu início há oito ou mais anos). Já 64% tiveram a primeira aplicação OLAP disponibilizada aos usuários da organização há no máximo cinco anos.

### 3. 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% identificaram esse envolvimento e o restante afirmou que este fato raramente acontece (30%), nunca acontece (10%) ou não podem afirmar (23%), indicando assim uma ausência de conscientização da importância da qualidade dos dados.



**Figura 1. (a) Identificam pessoas responsáveis (coordenador/executor) por ações de melhoria na qualidade dos dados transacionais; (b) Identificam pessoas pertencentes às áreas de negócio da organização envolvidas com ações de melhoria na qualidade dos dados**

Porém observa-se que 50% dos respondentes conseguiram identificar na organização coordenadores e/ou executores responsáveis por ações de melhoria de qualidade dos dados transacionais (Figura 1a), mas apenas 42% dos respondentes

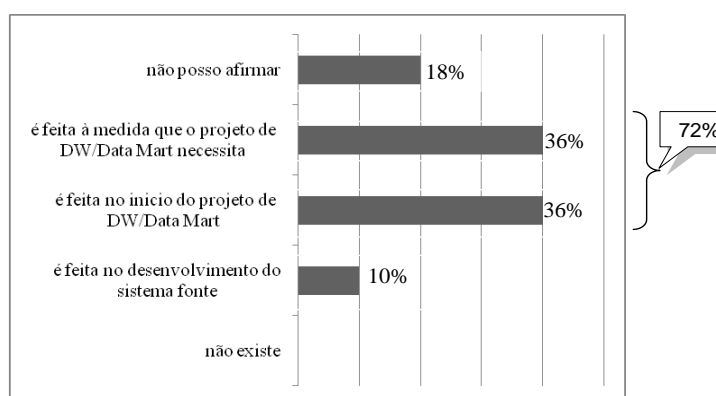
conseguiram identificar profissionais das áreas de negócio da organização envolvidos em ações de melhoria na qualidade dos dados, o que leva a crer que este assunto ainda não é bem tratado na organização, tendo ainda o seu enfoque restrito ao âmbito da TI (Figura 1b). A baixa identificação dos responsáveis pelas ações de melhoria na qualidade de dados e em especial de pessoas envolvidas nas áreas de negócio ilustra uma má organização das ações de melhoria na qualidade de dados e um enfoque de tais iniciativas na própria TI.

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% reconheceram que a utilização de dados com qualidade permitirá que a organização seja mais eficiente. Já 80% acharam que ela se torna mais organizada, 70% perceberam que ela se tornará mais ágil (contra 5% que acredita que ela se tornará mais burocrática) e 65% que se tornará mais competitiva, sugerindo assim um bom nível de conscientização do profissional de TI acerca dos ganhos efetivos obtidos na organização. 35% acreditam ainda que a organização se tornará mais lucrativa.

No que se refere a projetos de DW e/ou DM, cerca de 65% dos pesquisados afirmaram que a missão e os objetivos do DW e/ou DM encontram-se bem definidos (18% afirmaram que quase sempre, 46% afirmaram que com frequência, 27% afirmaram raramente e 9% que não podem afirmar).

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 and Dayal 1996], persistem dúvidas, indefinições e desconhecimento acerca dos efetivos objetivos deste projeto.



**Figura 2. Documentação das fontes de dados do projeto de DW e/ou DM**

Cerca de 70% afirmaram 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 2). 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/DWs, possivelmente nem todos os dados (ou 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.

Ainda se observa que 50% dos pesquisados afirmaram 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 afirmaram existir um processo que avalia a qualidade de dados existentes na organização, embora 50% dos pesquisados indicassem 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.

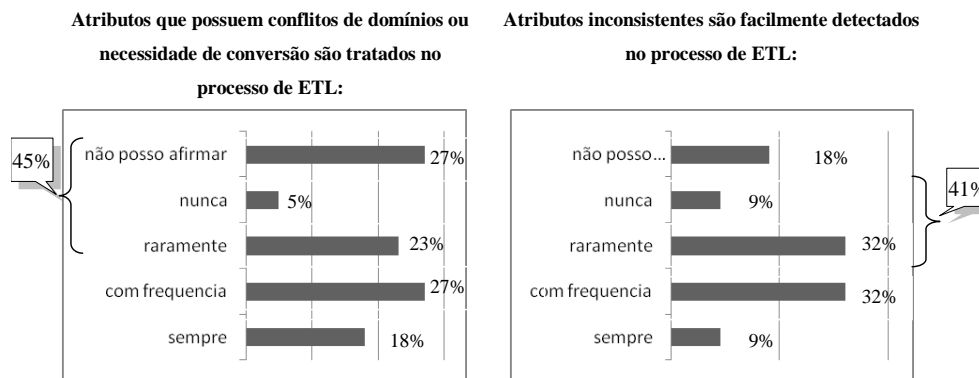
Ainda cerca de 60% dos respondentes afirmaram que existe (27% com frequência e 32% raramente) um repositório central de metadados, utilizado para documentar essas fontes de dados, provavelmente fornecidas pelas ferramentas utilizadas. O fato de 73% dos respondentes indicarem raramente (32%), não posso afirmar (32%) ou nunca (9%), 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.

Um dos respondentes afirmou que geralmente as ferramentas de ETL possuem um repositório central de metadados, considerado 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 acrescentou 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. A adoção de uma metodologia tende a possibilitar uma melhor gerência da qualidade dos dados na organização. Na pergunta específica para verificar a utilização de alguma metodologia, foram destacadas três (que em geral são as mais utilizadas): ISO 9001 [ISO 2008], GQM [Basili *et al.* 1994] e TDQM [Wang 1998]. 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 (9% utilizam TDQM e 9% a ISO 9001). Um resumo sobre metodologias para qualidade de dados pode ser encontrado em [dos Santos 2010]. Considerando as diversas propostas de metodologia, bem como a diversidade de critérios de qualidade de dados, havia uma expectativa de adoção de alguma metodologia e uma dificuldade em se escolher a metodologia e critérios mais adequados à empresa. Os resultados indicam um desconhecimento da importância da

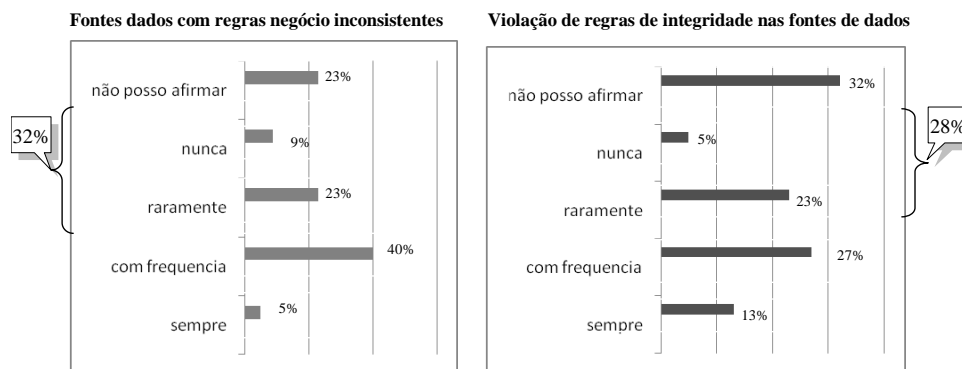
adoção de uma metodologia e a definição adequada de critérios, indicando que o problema vai além do caráter técnico.

Em seguida foram feitas algumas perguntas que se referiam aos tipos de tratamento de dados efetuados nas fases do processo de ETL e o quanto dos problemas encontrados eram devidamente tratados e registrados em um repositório de metadados (Figura 3). 28% dos respondentes indicaram que conflitos de domínios ou necessidade de conversão não são tratados (ou raramente) no processo de ETL (e outros 27% não podem afirmar). Além disto, 41% dos respondentes indicaram que atributos inconsistentes não são facilmente detectados no processo de ETL (outros 18% não podem afirmar). Isto indica claramente problemas de qualidade de dados.



**Figura 3. Tratamento dos atributos no processo de ETL**

Observa-se que em torno de 35% dos respondentes não conseguiram 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 4)



**Figura 4. Tratamento de regras de negócio no processo de ETL**

Cerca de 65% dos pesquisados reconheceram 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 respondentes acrescentaram comentários afirmando que na prática isto não acontece (Figura 5). 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 acrescentou que com isso o próprio usuário do BI comunica à equipe quando identifica um problema.

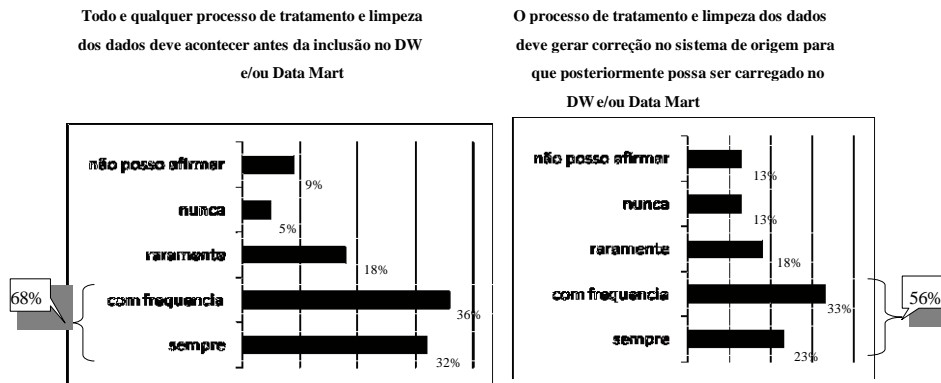


Figura 5. Avaliação do tratamento de dados efetuado no processo de ETL

Em torno de 50% dos respondentes afirmaram que regras de sumarização dos dados encontram-se identificadas e mapeadas em um repositório de metadados e 40% afirmaram 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 6).

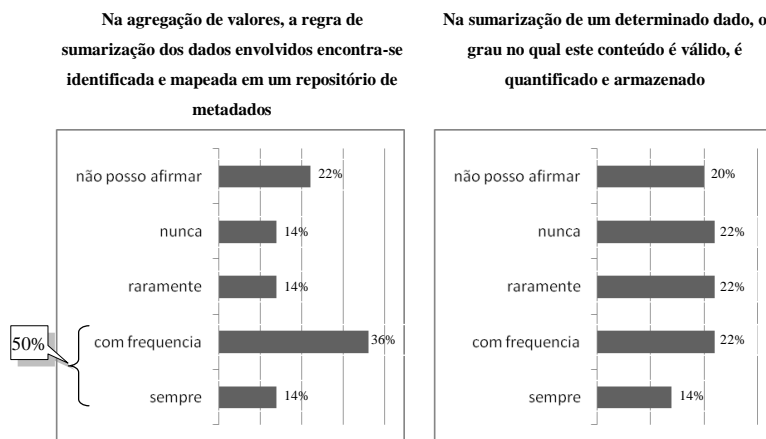


Figura 6. Avaliação do tratamento de dados na sumarização

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 7).

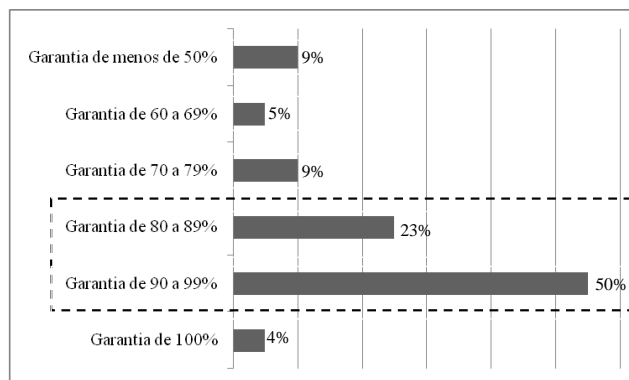
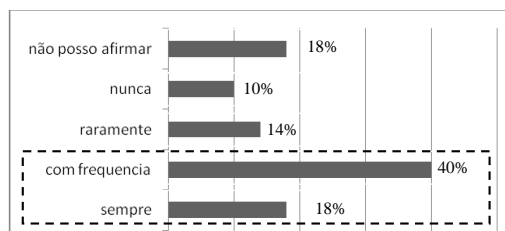


Figura 7. Avaliação da garantia da qualidade de dados no DW e/ou Data Mart executada pelo processo de tratamento e limpeza dos dados



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.

Formal ou informalmente, os profissionais de TI se mantêm atualizados sobre a satisfação dos usuários do DW e/ou DM, porém cerca de 60% afirmaram que os usuários continuam manifestando a necessidade de ter mais informações sobre a qualidade do conteúdo do mesmo (Figura 8).



**Figura 8. Quando um usuário manifesta necessidade de mais informações sobre a qualidade do conteúdo**

Finalmente, cerca de 60% dos pesquisados consideraram 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. 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.

#### 4. Conclusões

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.

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 em consideração 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 todas as áreas gestoras de informação na 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.

Alguns trabalhos futuros são: (i) uma investigação da qualidade de dados nas organizações sob o enfoque de apoio a decisão em um universo maior de empresas e profissionais; (ii) um estudo sobre a relação de causa e efeito entre a qualidade dos dados utilizados na tomada de decisões e a qualidade das decisões que são tomadas no ambiente organizacional; e (iii) a definição e disseminação de um programa de Governança de Dados capaz de 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. Este terceiro item é objeto de estudo atual dos autores.

## References

- Andrade, M. M. (2008), Como Preparar Trabalhos para Cursos de Pós Graduação, São Paulo: Atlas, 7ª edição.
- Basili, V.R., Caldiera, G. and Rombach, H.D. (1994) “Goal Question Metric paradigm”. Encyclopedia of Software Engineering, pp. 528-532, John Wiley & Sons.
- Campos, A. L. de S. (2005) “Indicadores de Qualidade de Dados em Data Warehouse”, Monografia de conclusão de curso. UFJF, Juiz de Fora, 48f.
- Chaudhuri, S. and Dayal, U. (1996) “Requirements capture and analysis: a survey of current practice”, Requirements engineering, v. 1, n. 2, pp. 75-87. Springer-Verlag.
- dos Santos, I.M.F. (2010) “Uma proposta de governança de dados baseada em um método de desenvolvimento de arquitetura empresarial”, Dissertação de Mestrado. UNIRIO, Rio de Janeiro, 140f.
- Friedman, T. and 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. (2002), Como elaborar projetos de pesquisa, São Paulo:Atlas, 4ª edição.
- International Organization for Standardization (2008): ISO 9001:2008: Quality Management Systems – Requirements. Beuth, Berlin.
- Vergara, S. C. (2009), Projetos e relatórios de pesquisa em Administração, São Paulo: Atlas, 10ª edição.
- Wang, R. (1998) “A product perspective on total data quality management”. Communications of the ACM, v. 41, n. 2, pp. 58-65.

## APÊNDICE 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 ...    mais de 10 anos
5. Há quanto tempo atua na área de TI?  
 menos de 1 ano    1 ano    2 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 ...    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 ...    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 ...    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

Frequência: 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 ...

1. 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
2. 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
3. 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
4. A documentação dos modelos de dados é mantida atualizada pela equipe de TI.  
 sempre    com frequência    raramente    nunca    não posso afirmar
5. Existe um processo que avalia a qualidade dos dados existentes na organização.  
 sempre    com frequência    raramente    nunca    não posso afirmar
6. 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
7. A utilização de dados com qualidade permitirá que a sua organização seja:  
 mais ágil    mais burocrática  
 mais competitiva    menos competitiva  
 mais lucrativa    menos lucrativa  
 mais eficiente    menos eficiente  
 mais organizada    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 apoiem este processo.  
 Discordo  Discordo parcialmente  Concordo parcialmente  Concordo totalmente