

Proposta de medição da qualidade da informação

Fábio Favaretto (PUCPR) fabio.favaretto@pucpr.br

Resumo

A multiplicação de fontes de dados e informações tem ampliado as possibilidades de acesso aos mesmos. Estas fontes podem ser internas ou externas à organização. A medição da qualidade da informação se torna uma forma de classificar e identificar as informações que podem atender melhor aos interesses e necessidades dos usuários. O objetivo deste artigo é apresentar uma proposta para a medição da qualidade da informação (QI) do ponto de vista do usuário. Esta proposta consiste de um processo de medição da QI, que pode ser utilizado nos fluxos de informações que ocorrem no suporte à decisões.

Palavras-chave: Qualidade da informação, qualidade de dados, medição da qualidade da informação.

1. Introdução

A multiplicação de fontes tem ampliado as possibilidades de acesso a um volume cada vez maior de dados e informações. Estas fontes podem ser internas ou externas à organização. Isso pode ser observado quando é feita uma consulta à Internet através de mecanismos de busca comum, onde são obtidos geralmente milhares de resultados, porém somente poucos destes atendem realmente as necessidades do usuário. Com o aumento da disponibilidade de informações, a medição da qualidade se torna uma forma de classificar e identificar as informações que podem atender melhor aos interesses e necessidades dos usuários.

A qualidade da informação tem sido vista como a apropriação da informação para o uso pelo consumidor (WANG *et alli*, 2000). Usualmente *dados* e *informações* são usados como sinônimos. É aceito que o processamento de dados gera informações. Como este artigo irá tratar da qualidade da informação (QI), entende-se que os dados também devem ter qualidade. Assim, o termo QI se refere tanto à qualidade das informações como dos dados.

Segundo Burgess *et alli* (2004), devido ao grande volume de informações disponível atualmente, a qualidade se tornou um importante fator de escolha entre qual informação utilizar e qual descartar. Lee e Siau (2001) também afirmam que diante da enorme quantidade de dados disponíveis são necessárias ferramentas que gerem informações úteis.

A estrutura deste trabalho é apresentada a seguir. Na próxima seção será apresentada a metodologia utilizada. A seguir é apresentado o conceito de qualidade da informação e são feitas considerações sobre a sua medição. Após isso será apresentada a proposta de um processo de medição da QI. Na seqüência são apresentados os resultados e as conclusões obtidas.

2. Metodologia

Este trabalho tem caráter exploratório, e o resultado esperado é a proposta de um processo de medição da qualidade da informação e uma referência para indicar as etapas do ciclo de geração e utilização da informação onde esta medição pode ser feita.

Caballero *et alli* (2004) afirmam que muito tem sido dito sobre a necessidade de qualidade da informação, porém poucos trabalhos apresentam considerações práticas sobre a garantia da QI. O objetivo geral do presente trabalho é explorar e operacionalizar a medição da QI, permitindo um controle mais acurado para sua garantia.

Em uma primeira etapa, serão definidas as possibilidades de medição da QI em um SI. Na etapa posterior será proposto um processo para a medição da QI, com foco nas possibilidades levantadas anteriormente.

3. Conceitos de QI

Uma pobre qualidade da informação tem severos impactos na efetividade geral de uma organização. Pesquisas revelaram que nos Estados Unidos, mais de 60% das firmas de médio porte com vendas anuais de mais de 20 milhões de dólares tinham problemas com qualidade da informação. Nos arquivos de registros criminais nos EUA 50% a 80% das informações são incompletas, ambíguas, ou sem acuracidade (WAND e WANG, 1996).

Segundo Khan *et alli* (2002), enquanto a adequação ao uso captura a essência da qualidade, a sua medição é difícil quando se utiliza este conceito geral. A visão da qualidade como conformidade com as especificações facilita sua medição.

Um conceito fundamental para se apresentar a QI é o de produto de informação (PI), que é uma analogia direta com produtos resultantes de processos de manufatura tradicionais. Estes produtos “físicos” são os resultados de um processo de produção onde matérias primas são processadas. O produto de informação é o resultado do processamento de dados (brutos) em sistemas de informação. Essa analogia pode ser vista na Tabela 1.

	Manufatura de produtos	Manufatura de informação
Entrada	Matérias primas	Dados brutos
Processo	Processo de manufatura	Sistema de informação
Saída	Produto “físico”	Produto de informação

Tabela 1: Analogia entre sistemas de manufatura de produtos e de informações (adaptado de WANG *et alli*, 2000).

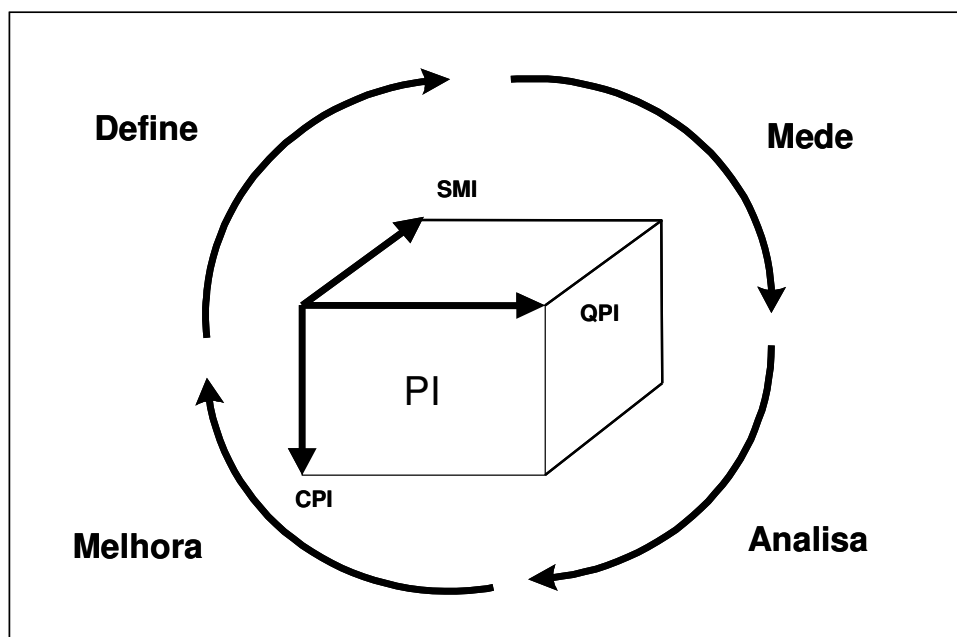
De acordo com Helfert e Heinrich (2003), a qualidade da informação pode ser diferenciada em duas, referentes às decorrências do projeto e da conformidade. Isso leva à informações que atendem as necessidades do usuário e ao seu uso. Mahnic e Rozanc (2001) colocam que a qualidade da informação é um pré-requisito para implantação e desenvolvimento de sistemas de informação, como por exemplo, o *Data Warehouse*, pois sem este atributo as informações resultantes deste sistemas também serão de baixa qualidade.

Assim como os produtos “físicos”, a informação possui dimensões da qualidade associadas. Essas dimensões são agrupadas em categorias, conforme pode ser visto na Tabela 2. O senso comum aponta a *acuracidade* como sendo a própria QI, porém podemos ver que esta é apenas uma das dimensões. A categoria *intrínseca* mostra o fato de a informação ter qualidade em si mesma. *Acessibilidade* e *representação* enfatizam a importância do papel dos sistemas de informação (SI), que fazem o papel do sistema de manufatura da informação (SMI). A categoria *contextual* mostra a inserção da informação em seu meio de utilização.

Categoria	Dimensão
Intrínseca	Acuracidade, objetividade, credibilidade e reputação.
Acessibilidade	Acessibilidade e segurança no acesso.
Contextual	Relevância, valor agregado, temporalidade, integridade e volume de informação apropriado.
Representação	Interpretabilidade, facilidade de entendimento, representação concisa, representação consistente e facilidade de manipulação.

Tabela 2: Categorias e dimensões da Qualidade da Informação (adaptado de WANG *et alli*, 2000).

A seguir é apresentado o modelo para a melhoria da qualidade da informação (TDQM - *Total Data Quality Management*). As tarefas são realizadas de maneira interativa. Uma visão deste modelo é apresentada na Figura 1.

Figura 1: Modelo para a melhoria da qualidade da informação (adaptado de Wang *et alli*, 2000).

Em sua aplicação, devem ser definidas as características do PI (CPI), definir as necessidades de qualidade para o PI (QPI) e identificar o sistema de manufatura da informação (SMI). Após as definições é possível passar para as medições, que irão fornecer a base para as análises. Estas devem investigar a causa raiz (principal) que pode gerar os problemas de QI. Os métodos de análise podem ser simples ou complexos, indo desde um controle estatístico do processo (CEP), análises de gráficos de Pareto e outros. Estas análises irão indicar quais bancos de dados ou funções do SMI devem ser escolhidos para maximizar a melhoria da QI.

4. Medição da QI

Neste trabalho, a medição da QI será definida como a atribuição de valores às dimensões apresentadas, segundo o ponto de vista do usuário do PI. Um sistema de informação (SI) pode ser visto como um processo de produção de produtos de PI. Assim, a medição da QI deve ocorrer no SI. Na Figura 2 são apresentados os elementos de um SI e seu relacionamento com o ciclo de decisão.

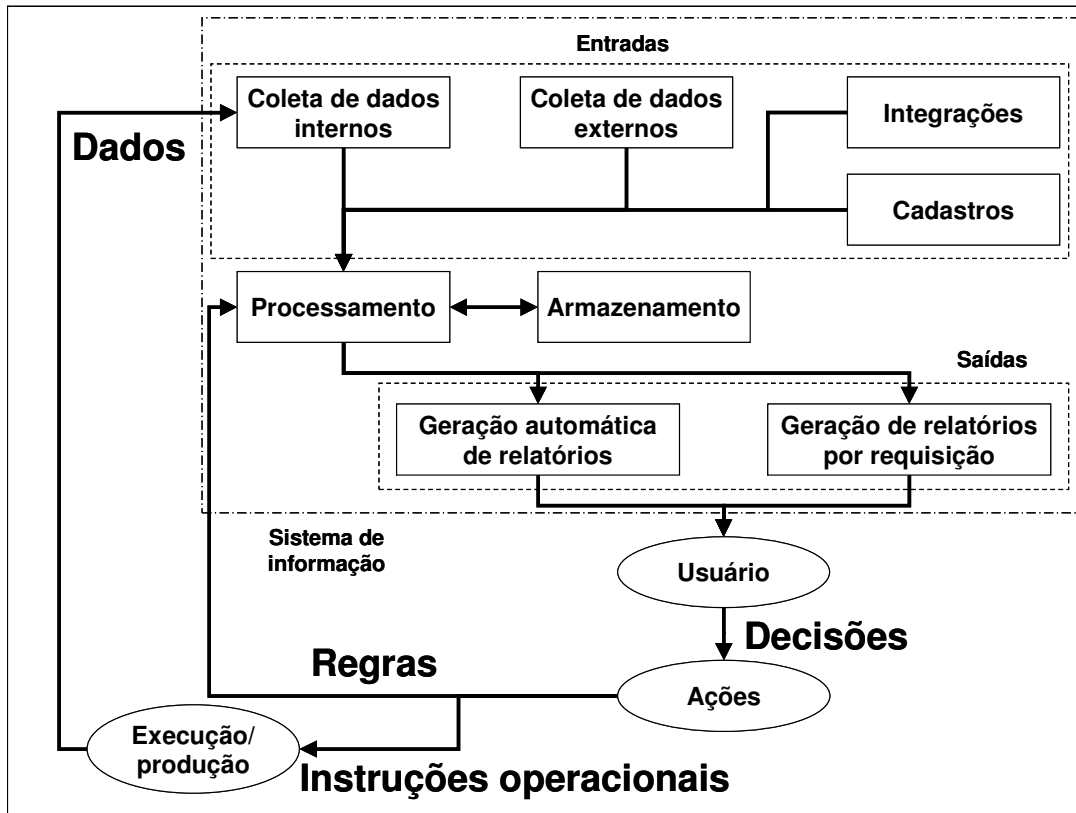


Figura 2: Ciclo de decisão utilizando um Sistema de Informação.

Os SI recebem entradas, as processam, armazenam e geram saídas. As possibilidades de melhoria da qualidade da informação estão nestes elementos.

As entradas dos SI podem ser divididas em quatro elementos. O primeiro deles é a *coleta de dados internos*, onde são feitas entradas de dados relacionados ao processo gerenciado pelo SI, e que na sua maioria são relacionados às atividades de execução ou produção. Outro elemento é a *coleta de dados externos*, que são aqueles relacionados ao ambiente, como a situação geral de um determinado setor de negócios. O resultado das *integrações* é outro elemento das entradas de um SI, representando todos os dados que são obtidos de outros SI, através de integrações entre bancos de dados, utilizando arquivos texto, EDI (*Electronic Data Interchange*), XML (*Extensible Mark-up Language*) ou ODBC (*Open Database Conectivity*). O último elemento de entrada de um SI é o conjunto de *cadastros* realizados no próprio aplicativo.

Nos elementos de entrada do SI todas as dimensões da QI podem ser analisadas. Entretanto, a categoria intrínseca (dimensões: acuracidade, objetividade, credibilidade e reputação) é aquela

que possui mais oportunidades para melhoria da QI, visto que as outras dimensões podem ser trabalhadas posteriormente.

O *processamento* é um o elemento do SI responsável pelas transformações dos dados e geração das saídas. Todas as dimensões de QI são envolvidas.

O elemento de *armazenamento*, apesar de não realizar transformações é responsável por algumas oportunidades de melhoria de QI, principalmente relacionadas às categorias: intrínseca e representação.

São dois os elementos de saída de um SI. O primeiro deles é a *geração automática de relatórios*, relacionado principalmente ao suporte de processos rotineiros e operacionais. O segundo elemento é a *geração de relatórios por requisição*, relacionado ao suporte de processos estratégicos e de decisões de alto nível, pois os relatórios são configurados pelo usuário e requisitados somente quando existe a necessidade. Estes elementos estão em contato direto com o usuário do SI e conseqüentemente do PI, assim, todas as dimensões da QI são importantes.

As medições podem ser realizadas de forma automatizada, onde podem existir funções ou aplicativos específicos para medir a qualidade de um dado à medida que ele é gerado ou transformado. Dessa forma não é necessário acionar a medição, sendo que neste caso, a medição tem o objetivo de *monitorar* a QI. Por outro lado, algumas medições podem ocorrer em dados já existentes, de forma não automatizada, onde o usuário ou o aplicativo dispara uma medição, com o objetivo de *controlar* a QI.

5. Processo de medição

O objetivo do modelo do TDQM é a gestão total da qualidade dos dados. O foco deste artigo é uma das etapas do TDQM, a medição da QI. Nesta seção será feita uma proposta para o detalhamento de como esse processo deve ser realizado.

Na Figura 3 é apresentado o processo proposto para medição da QI.

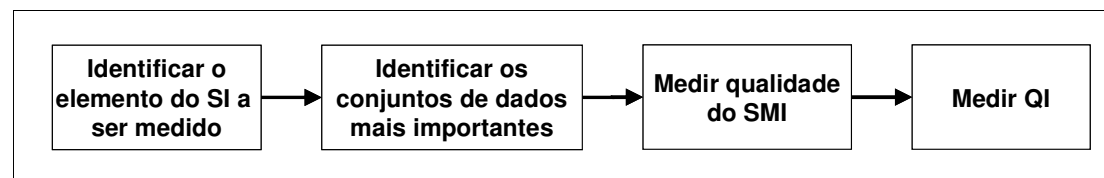


Figura 3: Processo para medição da QI.

A primeira atividade a ser realizada é a identificação do elemento a ser medido. Trata-se do elemento do Sistema de Manufatura da Informação (SMI), no caso o próprio Sistema de Informação (SI). Em função das necessidades do usuário, deve ser então identificado qual elemento do SI (conforme apresentado anteriormente na Figura 2) é mais crítico. Este elemento é composto, por sua vez, de blocos menores, que podem ser rotinas, módulos ou funções de um SI.

Dentro do elemento identificado (ou dos blocos que o compõem) devem ser identificados quais os conjuntos de dados são mais importantes. Estes conjuntos de dados podem ser partes fundamentais do PI ou serem importantes para os processamentos que irão gerar o PI.

Em relação aos conjuntos de dados mais importantes será feita uma medição da qualidade do SMI. Esta medição deverá verificar se o conjunto de dados está atendendo as necessidades do usuário e o propósito para qual foi gerado, sem considerar os aspectos de QI. Exemplificando, se o conjunto de dados mais importante for identificado como sendo um relatório dos produtos comercializados por uma empresa, podem ser medidas quantas vezes houve problemas na geração deste relatório. Esta medição não está relacionada diretamente com a QI, porém pode ser útil na etapa de análise do TDQM. Essa medição é feita com a adição de uma nova entidade ao modelo de dados (Modelo Entidade-Relacionamento – ME-R) que registra as ocorrências, e se relaciona diretamente com a(s) entidade(s) original(is) do modelo que contém os conjuntos de dados mais importantes. As ocorrências são aspectos que devem ser medidos para avaliar o correto funcionamento do SMI. O relacionamento desta nova entidade com a entidade original possui, por sua vez, um relacionamento com uma outra nova entidade que possui os valores para as ocorrências. No mesmo exemplo anterior, caso o conjunto de dados mais importante fosse um relatório dos produtos comercializados, haveria a criação de uma nova entidade que registra se a geração do relatório foi satisfatória, atribuindo valores em uma escala de satisfação. Estes valores de satisfação estariam armazenados em uma entidade específica para tal.

A medição da QI propriamente dita também é feita com o acréscimo de novos elementos ao modelo de dados do SMI. Segundo Kim *et alli* (2003), a medição e quantificação da qualidade dos dados devem ser desenvolvidas para serem utilizadas como guia na utilização e escolha de conjuntos de dados dentro de um ambiente com muitas opções. Os conjuntos de dados identificados anteriormente podem ser decompostos em conjuntos menores ou até mesmo em dados individuais. Para cada um destes deve ser feita a definição de quais dimensões da QI são relevantes e conseqüentemente serão medidas. Estas dimensões são armazenadas em uma nova entidade do modelo de dados, que se relaciona com a entidade original que registra o conjunto (ou o dado) em foco. Este relacionamento se relaciona com uma outra nova entidade que registra os valores para as dimensões de QI que são medidas.

A atribuição do valor de uma dimensão pode ser feita automaticamente, quando é possível identificar parâmetros que podem ser contados ou analisados pelo próprio SI. Por exemplo, na medição da dimensão integridade é possível contar automaticamente quantos registros estão completos (íntegros), fazer uma proporção dos registros incompletos com os registros completos e atribuir uma medida para a dimensão, em função de intervalos previamente definidos. Em outras situações, a atribuição deste valor não pode ser feita automaticamente, pois depende de uma avaliação (subjetiva ou objetiva) do usuário.

6. Conclusões

A medição da qualidade da informação acarreta no desenvolvimento de algumas atividades, que vão desde a definição das dimensões da QI, os critérios de medição e o projeto e implantação de novas funcionalidades no SI. Caso seja feito apenas o controle da qualidade do SI (aplicativo), o esforço necessário é consideravelmente menor, entretanto não se mede e conseqüentemente não se controla a QI. Para que isso seja feito é necessário um esforço maior.

A utilização de uma sistemática para medição da QI permite que isso ocorra de forma objetiva. A partir da identificação das possibilidades de medição e da realização das atividades do processo proposto é possível conhecer a QI e conseqüentemente a própria informação.

A melhoria da QI permite melhores entradas para as decisões, criando a chance de se obterem resultados melhores. De uma forma geral, as comunicações que ocorre em vários níveis dentro e fora da empresa também dependem da QI e podem ser melhorados com a melhoria desta.

Agradecimento

Para a realização deste trabalho o autor recebeu apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq.

Referências

BURGESS, M. S. E.; GRAY, W. A.; FIDDIAN, N. J. Quality measures and the information consumer. *Proceedings of the Ninth MIT International Conference on Information Quality*, Cambridge, Massachusetts, 2004.

CABALLERO, I.; GÓMEZ, O.; PIATTINI, M. Getting better information quality by assessing and improving information quality management. *Proceedings of the Ninth MIT International Conference on Information Quality*, Cambridge, Massachusetts, 2004.

HELFERT, M.; HEINRICH, B. Analyzing data quality investments in CRM: a model-based approach. *Proceedings of the Eighth MIT International Conference on Information Quality*, Cambridge, Massachusetts, 2003.

KHAN, B. K.; STRONG, D. M.; WANG, R. Y. Information quality benchmarks: product and service performance. *Communications of the ACM*, Vol. 45, N. 4, 2002.

KIM, W.; CHOI, B.; HONG, E.; KIM, S.; LEE, D. A taxonomy of dirty data. *Data mining and knowledge discovering*, N. 7, 2003.

LEE, S. L.; SIAU, K. A review of data mining techniques. *Industrial Management and Data Systems*, Vol. 101, N. 1, 2001.

MAHNIC, V.; ROZANC, I. Data quality: a prerequisite for successful data warehouse implementation, *Proceedings of the Seventh International Conference of European University Information Systems*, Berlin Humboldt-University, 2001

WAND Yair & WANG Richard Y. (1996) - Anchoring Data Quality Dimensions in Ontological Foundations, *Communications of the ACM*, 39.

WANG, R. Y.; ZIAD, M.; LEE, Y. W. *Data quality*. Kluwer Academic Publishers, 2000.

WANG, R. Y.; REDDY, M. P.; KON, H. B. Toward quality data: an attribute-based approach. *Decision Support Systems*, Vol. 13, 1995.