

QUALIDADE DA INFORMAÇÃO: UM ESTUDO SOBRE TÉCNICAS, FERRAMENTAS E PRÁTICAS PARA DESENVOLVIMENTO DE SISTEMAS DE INFORMAÇÃO

Renato Virto MOREIRA, Osmar
Aparecido MACHADO

Introdução

O presente trabalho trata-se de estudo sobre o uso das técnicas, ferramentas e práticas utilizadas para avaliar a qualidade das informações nos processos de desenvolvimento de software. Realiza uma investigação sobre o uso dos conceitos da qualidade da informação e, ao mesmo tempo, contribui para ampliar o escopo de conhecimentos sobre a qualidade da informação, especialmente orientado para as empresas e profissionais da área de desenvolvimento de software.

Parece existir um distanciamento entre o conhecimento científico e as práticas adotadas pelo mercado. É neste cenário que este estudo justifica sua viabilidade. Uma vez identificadas as melhores práticas de gestão da informação utilizadas pelo mercado e associadas com os conhecimentos e o rigor científico, o projeto poderá trazer

contribuições significativas também para o ensino dessas técnicas e ferramentas nas disciplinas relacionadas ao desenvolvimento de sistemas de informação.

Mesmo sendo a Qualidade da Informação (QI) uma área já consolidada, faltam iniciativas no sentido de popularização das técnicas e uso dos conceitos relacionados ao assunto. Embasado nos achados de Machado (2013), dentre outros, presume-se que tanto as empresas desenvolvedoras de software quanto os profissionais de tecnologia da informação, em sua maioria, não utilizam os conceitos de *data quality* em seus processos de desenvolvimento de software.

Para tanto, foi realizada uma extensa pesquisa bibliográfica sobre o assunto, que segundo Gil (1993, p. 48), esse tipo de pesquisa é “desenvolvida a partir de material já elaborado, constituído principalmente de livros e artigos científicos. Embora em quase todos os estudos sejam exigido algum tipo de natureza, há pesquisas desenvolvidas exclusivamente a partir de fontes bibliográficas”.

Relevância da Qualidade da Informação

A informação tem passado por diversas transformações ao longo do tempo e cada uma dessas transformações reflete a interpretação do seu significado em cada momento. Numa extensa revisão do conceito, Capurro e Hjørland (2007, p. 154-156) resgataram algumas reflexões e definições em que as transformações são percebidas e sobre o uso do termo informação, relatam que dever-se-ia ter em mente que “informação é o que é informativo para uma determinada pessoa” e o “que é informativo depende das necessidades interpretativas e habilidades do indivíduo”. Ressaltam que essa é uma visão mais atual do termo e que o sentido de uso da palavra, de origem latina, tem se alterado significativamente através dos tempos: *informatio* e *informo* eram inicialmente utilizados no sentido de dar forma a algo ou de moldar a mente. Outra interpretação do termo refere-se às formas de entendimento de como a mente era informada pelas sensações do mundo: “parecia descrever os mecanismos da sensação: os objetos do mundo in-formam os sentidos [...]”; Entretanto, ressaltam os autores que “a sensação é inteiramente diferente da forma, uma é sensorial, a outra

intelectual; uma é subjetiva, a outra objetiva”.

Dentre tantas outras interpretações, Machado (2013, p23) salienta que a busca pelo entendimento da informação é complexa, pois ora ela “é descrita como um objeto ou serviço, ora como uma ideia, um pensamento” e as dificuldades de chegar ao consenso sobre o que de fato é informação “perpassam por várias épocas e áreas do conhecimento, cada qual procurando interpretá-la à luz do desenvolvimento social, tecnológico ou cognitivo disponível”. Nesse sentido, entender a informação e explorar suas características têm sido o objeto de inúmeros estudos e a diversidade da produção científica sobre o assunto contribuiu para o entendimento atual da informação, sobretudo em relação aos aspectos sociais e cognitivos.

A partir do início da chamada era da informação, em meados da década de 60, o termo passou a assumir um protagonismo crescente, ampliando-se exponencialmente a partir do surgimento da internet, nos anos 90. Isto porque até os anos 60/80 o seu registro digitalizado estava devidamente organizado e controlado em bancos de dados,

relatórios e anotações, conforme as limitações da época. A gestão da informação, ainda incipiente, ocupava-se basicamente de garantir formas adequadas de armazenagem e recuperação da informação, especificamente nas organizações. Aproximadamente 40 anos depois, o termo informação reina absoluto em todos os segmentos da sociedade.

Já não se discute na atualidade sobre a importância da informação para as organizações, mas sim sobre a sua adequação ao uso. Apesar dos investimentos realizados para obtê-las, as informações não são ainda utilizadas de forma consistente pelos gestores e, por consequência, impactam financeiramente ou não, na obtenção de vantagens pelo seu uso.

E, nesse sentido, sabe-se que informações de qualidade tendem a propiciar melhores resultados para quem as utiliza. Machado (2013) lembra, no entanto, que identificar e mensurar os atributos que possibilitam avaliar a qualidade da informação não é tarefa fácil. O produto informação carrega características intrínsecas de poder ser utilizada e atender às várias áreas do conhecimento e nos contextos mais diversos. Repo (1989) salienta que a informação é humana, somente há

informação por intermédio da observação humana. Ela é multiplicável, quanto mais a usamos, mais útil ela se torna; é substituível, pode substituir outros recursos como dinheiro, pessoas, matéria-prima; é transferível, a velocidade e a facilidade com que se transmite uma informação é um fator a ser considerado para o desenvolvimento das demais áreas do conhecimento; é difusiva, tende a se tornar pública, mesmo que não seja o desejo de quem a detém; é compartilhável, ao contrário de outros bens, quem a vende ainda permanece com ela.

Existem técnicas, conceitos e ferramentas, como as desenvolvidas e apresentadas por Wang e Strong (1996), Strong e Wang (1997) e Wang (1998), que possibilitam melhor gestão das informações no sentido de potencializar o seu uso e sua qualidade para as organizações. Falta, contudo, maior popularização dessas técnicas e conceitos, para que sejam adotadas e implementadas para que promovam as boas práticas de gestão da informação. Machado (2018) ressalta a importância de estudos que tratam do fomento ao uso da qualidade da informação e da sua contribuição tanto para a disseminação desses conceitos no ambiente acadêmico quanto para as organizações, no sentido

de experimentarem técnicas que as ajudem a maximizar os seus resultados.

Os estudos, em nível mundial, sobre a qualidade da informação demonstram bons resultados acerca da relevância da qualidade da informação. No Brasil, entretanto, existe carência de estudos relacionados ao tema e especificamente na área de desenvolvimento de softwares esses conceitos são essenciais, pois é por meio dos aplicativos que a maioria das informações são introduzidas e registradas nas organizações. Dessa forma, objetivando contribuir e ampliar o conjunto de conhecimentos sobre a qualidade da informação, é que esse estudo buscará identificar e descrever conceitos, técnicas e ferramentas que tratam do uso da qualidade da informação nos processos de desenvolvimento de software.

2.1. Qualidade da informação

Qualidade da informação e qualidade de dados são tratados geralmente como sinônimos na literatura, embora existam diferenças conceituais entre os termos, conforme distingue Machado (2018):

- **Dados:** São elementos por meio dos quais as informações são geradas. São elementos brutos,

descontextualizados, cuja interpretação provê as informações necessárias.

- **Informação:** É a interpretação de um ou mais dados, que contextualizada expressa significado, capaz de influenciar e orientar para uma decisão.

Dessa forma, neste estudo, optou-se por adotar o termo Qualidade de dados quando se fizer referência exclusivamente ao tratamento dos dados; e Qualidade da Informação quando tratar-se da informação tratada, classificada e contextualizada.

O registro de dados de baixa qualidade gera diversos problemas, que na maioria dos casos são conhecidos na área de desenvolvimento de Software. Para justificar uma melhoria na qualidade de dados, nas empresas, existem diversos processos de medição de qualidade de dados (QD) que são necessários para obter correção e ação preventiva sobre esses dados. Para que tais processos transcorram são necessários investimentos significativos de tempo e dinheiro, o que torna a técnica onerosa, além disso as correções desses dados talvez não sejam perceptíveis facilmente, o que geralmente dificulta a valorização da mesma.

Entretanto, a melhoria da Qualidade de Dados traz novas oportunidades para a

empresa, como tomadas de decisões que impedem a perda de dinheiro em um falso investimento, ou ganhe dinheiro com bons investimentos. Infelizmente não são todas as empresas que investem para melhorar a qualidade dos dados, normalmente estas quando encontram dados inadequados, realizam investimentos suficientes para corrigir os erros temporários, e não investimentos para corrigir erros futuros, (BOBROWSKI; SOLER, 2007).

Segundo Salvador et al. (2006), a qualidade de dados pode ser definida como “adequação ao uso”, o que implica ser este um conceito relativo, uma vez que dados considerados de qualidade apropriada para determinado uso podem não ter qualidade suficiente para outro.

Entenda-se também por qualidade de dados o grau de aderência entre as visões apresentadas pelos sistemas de informação e os mesmos dados no mundo real. Um sistema com qualidade de dados de 100% indicaria perfeita aderência com o mundo real, enquanto uma qualidade de dados de 0% constituiria total falta de aderência (SALVADOR et al., 2006).

Para Wang *et al.* (2000), a qualidade da informação é vista como a apropriação

da informação para o uso pelo consumidor. Existem tecnologias para o armazenamento de dados que realizam alguma forma de tratamento e que melhoram a qualidade de dados por meio de técnicas como mineração de dados, como *Data Warehouses*, *Data Marts* e o próprio *Data Mining* (SALVADOR, et al., 2006). Na maioria dos casos, essas tecnologias estão alinhadas com os objetivos táticos e estratégicos das organizações, porém, as fontes que disponibilizam os dados que alimentam esses repositórios são os bancos de dados relacionais, ou seja, que coletam e armazenam a maioria dos dados no nível operacional, do dia-a-dia das organizações, por meio de sistemas tradicionais que efetuam pagamentos, recebimentos, vendas, dentre outros.

Por isso é fundamental que as informações extraídas e disponibilizadas por esses sistemas sejam de boa qualidade. E, como observa Salvador et al. (2006), na realidade, pouco se sabe da qualidade dessas informações, se realmente são de qualidade. Elas são disponibilizadas pelas aplicações (softwares) e começam a fazer parte das estatísticas das organizações. O objetivo não é que uma empresa tenha qualidade em 100% de suas ações, mas que os dados sejam precisos, atualizados e

consistentes suficientemente para que não afetem os processos de tomadas de decisão.

Técnicas e modelos de gerenciamento da Qualidade

As técnicas de medição da qualidade de dados são essenciais para avaliar os dados armazenados em bancos de dados, se são precisos e úteis para estatísticas empresarias ou tomadas de decisão.

Durante a realização desse estudo, observou-se a existência de várias iniciativas desenvolvidas ao longo dos últimos anos sobre modelos, técnicas e ferramentas para melhorar a experiência da qualidade da informação nas organizações. Após revisão da literatura e análise de diversas propostas, optou-se, nesse estudo, por avaliar e apresentar algumas técnicas e modelos possíveis de serem implementados e utilizados tanto em âmbito acadêmico, para o ensino da qualidade de dados, quanto pelas organizações a fim de colocar e experimentar melhores técnicas e práticas para tratar a qualidade de dados e da informação.

3.1. Gerenciamento da Qualidade de Dados

A DQM - *Data Quality Manager* (Gerenciamento da Qualidade de Dados)

é uma proposta elaborada por Angeles & Mackinnon (2005) e visa gerenciar a comunicação dos processos dos usuários com as aplicações a fim de resolver problemas de qualidade semântica dos dados. A técnica baseada em um modelo de referência (*Data Quality Reference Model*), cuja consideração principal está calcada em dois aspectos da qualidade: o foco interno e o foco externo.

Como salientam os autores, a qualidade de dados não se trata de um valor absoluto e cada consumidor de dados tem expectativas e necessidades diferentes em função de suas características, contexto e conhecimento. Por exemplo, de um lado os consumidores podem preferir alguns dados ao invés de outros devido à reputação dos produtores, à credibilidade e relevância dos dados para a tarefa, ou o nível de satisfação para a tomada de decisões estratégicas efetivamente usando dados confiáveis. De outro lado, usuários de nível operacional pode ter um interesse maior apenas na acessibilidade dos dados, a sua disponibilidade e sua pontualidade para realizar um processamento efetivo (ANGELES & MACKINNON, 2005).

3.2. Estimativa da Qualidade de Dados

Motro e Rakov, (1998) realizaram um

trabalho capaz de realizar estimativas sobre a qualidade de dado, estruturada em três pontos principais: proteção, medição e melhoria. O contexto da proposta são os bancos de dados relacionais. O modelo é formalizado sob duas instancias de um esquema de banco dados, uma considerando os dados armazenados no banco de dados e outra, o mundo real, em que:

D = Atual: Instancia D, denota o banco de dados (dados armazenados);

W = Ideal: Instancia que denota o mundo real, do banco de dados.

Motro e Rakov (1998) ressaltam que W é uma instancia hipotética e indisponível; e, a instancia D é uma aproximação da instancia ideal (W). Para determinar uma boa aproximação é preciso uma medida de similaridade das duas instancias.

Considerando que cada instancia é um conjunto de tuplas, pode se usar medidas para comparar os dois conjuntos de elementos, da seguinte forma.

- Estimativa de solidez (Soundness) = medida da proporção em que uma informação armazenada é verdadeira:

$$\frac{|D \cap W|}{|D|}$$

- Estimativa de Completude = medida da proporção em que uma informação verdadeira está armazenada.

$$\frac{|D \cap W|}{|W|}$$

Realizada a primeira etapa, denominada estimativa simples, outras duas etapas são executadas, uma para o refinamento da estimativa e uma terceira, que visa estimar a qualidade das respostas do processo. Ou seja, são realizados constantes desdobramentos em cada estimativa em que cada etapa carrega informações do anterior, gerando uma sequência de informações sobre os sucessivos refinamentos pelos quais passa o processo.

Por fim, por conta dos custos do processo, Motro e Rakov, (1998) ressaltam como uma limitação do método, que o mesmo seria mais adequado para informações estáticas. Já em situações em que as informações são dinâmicas, seria interessante registrar a qualidade no momento em que elas são obtidas. O estudo é uma contribuição importante e abre possibilidades para novas investigações e análises sobre o assunto.

3.3. Mediação total da Qualidade de Dados

Para uma melhoria do controle dos dados, Bertolazzi e Scannapieto (2001) entendem que cada organização social (na Itália) deveria disponibilizar seus dados para as outras organizações e vice versa, para que se tenha uma forma nova de enfrentar as barreiras tecnológicas e organizacionais.

Para tanto, propuseram um estudo sobre a qualidade de dados em sistemas cooperativos. Classificaram as dimensões da qualidade de dados em três categorias: visão conceitual, valores e formato. A proposta do estudo de Bertolazzi e Scannapieto (2001) foi embasada nessa classificação e realizada

por meio da metodologia *Total Data Quality Management* - TDQM, proposta por Wang (1998). As principais dimensões avaliadas foram selecionadas sob duas perspectivas: as dimensões intrínsecas (acurácia, completude, atualidade e consistência interna) que são próprias dos dados; e as dimensões específicas do processo, especificamente na troca de informações dentro do processo. Assim, consideraram a organização origem (A), a organização destino (C) e a troca realizada (B).

A figura 1 ilustra o intercâmbio de dados entre as organizações.

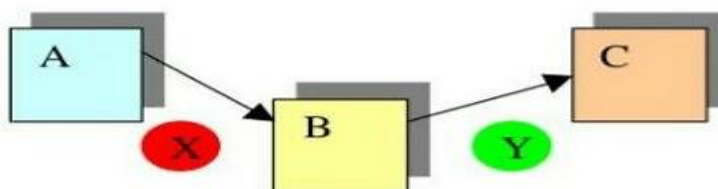


Figura 1- Trocas cooperativas entre três organizações diferentes

Fonte: Bertolazzi e Scannapieto (2001)

Consideraram ainda, as dimensões pontualidade (*timeliness*); a disponibilidade dos dados no processo, dentro das especificidades da organização destino; a importância: o significado do dado para a organização destino; e finalmente, a confiabilidade da fonte, que se refere à credibilidade da organização origem em relação ao

dado disponibilizado.

Uma das etapas fundamentais do modelo é a medição dos dados e como funciona sua estrutura de medição dentro do processo. O exemplo utilizado pelos autores é de uma base de dados, denominada REAL, com uma tabela chamada CITIZEN (que registra os dados dos cidadãos) é composta pelos

campos Nome, Sobrenome e SSN. Na sequência, o desenvolvimento de tabelas denominadas IDEAIS (sem erros), que serão utilizadas para estabelecer as métricas e comparações entre ambas. São compostas com os mesmos campos

da tabela REAL, mas com objetivos diferentes de medição (Figura 2). Essas tabelas são utilizadas dentro do processo para registrar a qualidade dos dados que estão sendo trocados e movimentados.

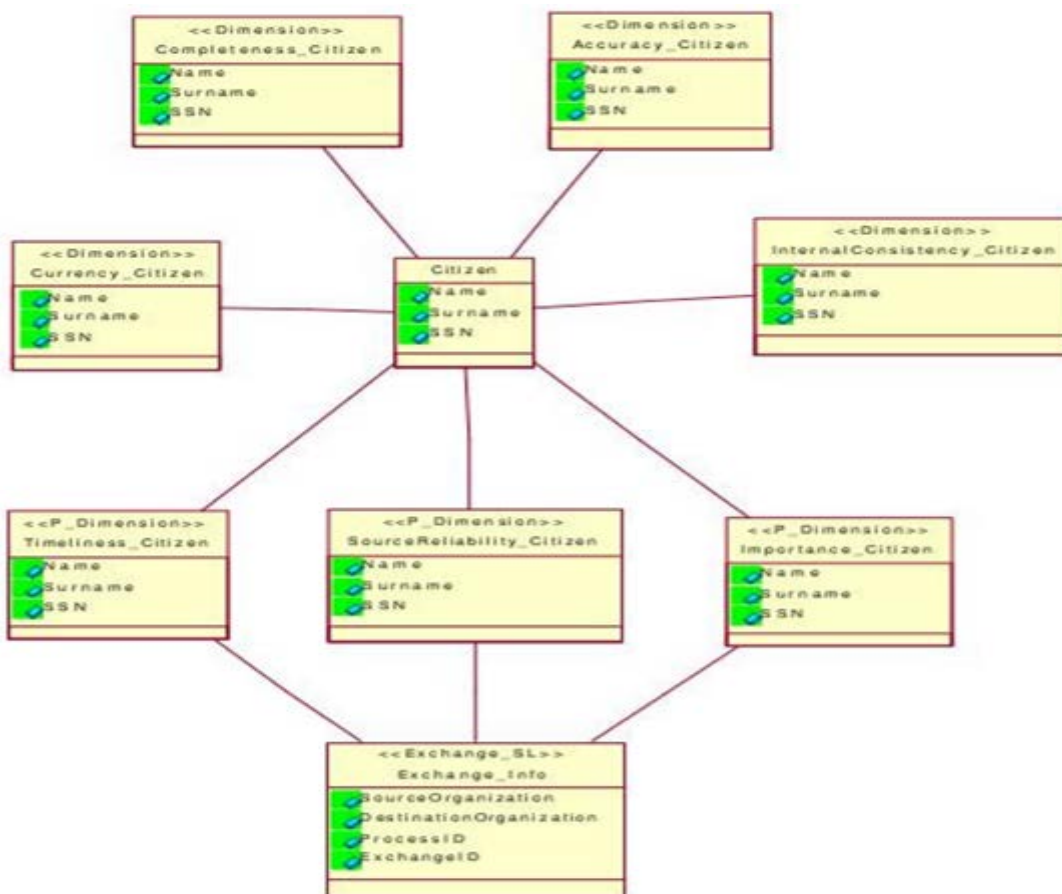


Figura 2 - Esquema cooperativo de qualidade de dados: Classe Cidadão

Fonte: Bertolazzi e Scannapieto (2001)

Em seguida e utilizando o ciclo de Gestão Total de Qualidade de Dados (TDQM), realizam-se as seguintes ações para implementação da proposta:

- Definição: Identificação das dimensões da qualidade de dados e dos requisitos relacionados;

- Medição: etapa que produz as métricas. Provê o feedback para o gerenciamento da qualidade e permite comparações entre a qualidade efetiva com os requisitos da qualidade pré-definidos.

- **Análise:** identifica as raízes do problema de qualidade e então, estuda seus relacionamentos.
- **Melhoria:** Promove a melhoria da qualidade por meio de essas técnicas.

3.3.1. TDQM cooperativa

Esta é uma etapa fundamental da proposta de Bertolazzi e Scannapieto (2001) para criação da TDQM cooperativa. Ela considera o ciclo produtivo da informação similar ao ciclo produtivo de qualquer produto, em que matérias primas entram no processo e saem como produtos acabados.

A proposta é implementada por meio de um conjunto de fases que se realizam de forma cíclica, sendo: Definição, mediação, exportação, análise e melhoria da qualidade.

Dentre os vários critérios específicos de cada uma das fases, é importante salientar a determinação de dois tipos de medição, que é uma das etapas fundamentais da proposta, na TDQM Cooperativa:

- **Estática:** neste tipo de medição, a confiabilidade dos dados e as dimensões intrínsecas são feitas estatisticamente, ou seja, cada

organização faz a medição de seus dados utilizando métodos tradicionais.

- **Método Dinâmico:** para este tipo de medição, apenas a pontualidade é medida dinamicamente, onde cada organização deve informar o tempo indicado para atualizar seus dados.

Os processos avaliados na proposta de Bertolazzi e Scannapieto (2001), corroboram para a implementação de técnicas que visam a qualidade e buscam disponibilizar dados a partir de critérios como: precisão, disponibilidade, relevância, completude, simplicidade e confiabilidade.

Conclusão

O aumento constante do uso de tecnologias e a necessidade das organizações por processos mais ágeis e de melhor qualidade, torna essencial que as empresas façam a gestão da qualidade de seus dados a fim de obterem melhores informações. Tais empresas, hoje ganham mercado através da análise dos dados, por exemplo, muitos bancos ganham dinheiro e mercado através da análise dos dados. O contrário pode gerar consequências incalculáveis. Justifica-se dessa forma, a necessidade da Qualidade dos Dados, abordado em tal projeto.

O objetivo deste estudo foi identificar métodos e técnicas para a melhoria da Qualidade dos Dados, este foi atingido a partir de uma ampla revisão bibliográfica. Feito isso, foi possível identificar a necessidade da medição da Qualidade dos Dados em várias áreas de atuação, tanto em empresas de tecnologia, como empresas automobilísticas, empresas comerciais e também na área da saúde.

Este trabalho pode contribuir para o desenvolvimento de futuros projetos de forma prática, utilizando este como uma base teórica para o aperfeiçoamento de tais técnicas. Há espaço e necessidade

para o desenvolvimento de ferramentas que utilizem essas técnicas e medidas abordadas neste projeto, tais ferramentas poderiam estar vinculadas aos Bancos de Dados, fazendo a aferição de cada campo, de cada tupla e informando aos desenvolvedores e gestores o nível de qualidade das informações utilizadas, se elas estão ou não adequadas para atender às necessidades da empresa.

E, por fim, necessário se faz informar que o presente trabalho foi realizado com o apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil.

Referências

ANGELES, Pilar; LACHLAN, Mackinnon,. Quality Measurement and Assessment Models Including Data Provenance to grade Data Sources. 2005. Disponível em: <https://www.semanticscholar.org/paper/Quality-Measurement-and-Assessment-Models-including-Angeles-Student/30adde50cc8b55171d80598fe052b8995a56c9bf>. Acesso em 18 jun 2018.

BERTOLAZZI, Paola. SCANNAPIECO, Monica. Introducing Data Quality in a cooperative context. 2001. Istituto di Analisi dei Sistemi ed Informatica Consiglio Nazionale delle Ricerche (IASI-CNR) Viale Manzoni 30, 00185 Roma, Italy. Disponível em: <<https://pdfs.semanticscholar.org/2d58/71dafec4c8c4c157ac6a2dd50eb6bf669f7c.pdf>>. Acesso em: 18 de julho de 2018.

BOBROWSKI, Monica; SOLER, Sabrina Vazquez. DQ Options: Evaluating Data -Quality Projects Using Real Options. 2007. Disponível em: <http://www.idea-group.com>. Acesso em 20 jun 2018.

CAPURRO, Rafael and HJØRLAND, Birger, O conceito de informação. Perspectivas em ciência da informação. [online]. 2007, vol.12, n.1, p. 148-207. Disponível em: <http://www.scielo.br/scielo.php?pid=S141399362007000100012&script=sci_arttext>. Acesso em: 10 jun. 2018.

CERVO, Amado Luiz; BERVIAN, Pedro Alcino; SILVA, Roberto da.

Metodologia Científica. 6 ed. São Paulo: Pearson Prentice Hall, 2007.

GIL, Antônio Carlos. Como Elaborar Projetos de Pesquisa. 3. ed. São Paulo, Atlas: 1993.

MACHADO, Osmar A. Qualidade da Informação: Uma abordagem orientada para o contexto. – São Paulo : Novas Edições Acadêmicas, 2018.

MOTRO A., RAKOV I. Estimating the quality of databases. In: Andreasen T., Christiansen H., Larsen H.L. (eds) Flexible Query Answering Systems. FQAS 1998. Lecture Notes in Computer Science, vol 1495. Springer, Berlin, Heidelberg, 1998. Disponível em: <https://link.springer.com/chapter/10.1007%2FBFb0056011> Acesso em: 05 Ago 2018.

OLETO, Ronaldo Ronan. Percepção da qualidade da Informação. Ciência da Informação. Brasília, v. 35, n.1, p. 57-82, jan./abr. 2006. Disponível em: www.scielo.br/pdf/ci/v35n1/v35n1a07.pdf. Acesso 29 jun. 2017.

REPO, A. J., The value of information: Approaches in economics, accounting, and management science. Journal of the American Society for Information Science and Technology, v. 40, p. 68–85, Mar. 1989. Disponível em: [http://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-571\(198903\)40:2%3C68::AID-ASI2%3E3.0.CO;2-J/abstract](http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-571(198903)40:2%3C68::AID-ASI2%3E3.0.CO;2-J/abstract). Acesso em: 15 Jun. 2017.

SALVADOR, Valéria Farinazzo Martins, et al. Qualidade de Dados para Gestão de Conhecimento na Área de Saúde. 2006.

STRONG, M. D; WANG, R. Y. Data Quality in Context. Communication of the ACM. May 1997, vol. 40, n. 5. Disponível em: http://www.fing.edu.uy/inco/cursos/cal_datos/articulos/StrongLeeWangCACMMay97.pdf. 15 Jun. 2017.

WANG. R. Y. and STRONG, D. M. Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information System, Spring 1996, v. 12, n.4, ABI/INFORM Global. Disponível em: [http://www.thespatiallab.org/resources/data%20quality\(JMIS\).pdf](http://www.thespatiallab.org/resources/data%20quality(JMIS).pdf). Acesso em: 20 Jun. 2017.

WANG. R. Y. A Product Perspective on Total Data Quality Management. Communications of the ACM. Feb./1998. v. 41. n.2. Disponível em: <http://dl.acm.org/citation.cfm?id=269022>. Acesso em: 20 Jun. 2017.

ZANETTI, G.C.C; MACHADO, O.A. *Implementação da Qualidade da Informação em Processos Gerenciais*. IX Fórum Científico da FEMA, 10,11,13,14 out. Assis, Fema, 2016. Disponível em: http://scorpion.femanet.com.br/forumcientifico/files/revistas/anais_ix_forum.pdf. Acesso em 15 Jun. 2017.