

Uma Análise das Relações entre a Qualidade da Informação e Big Data

Priscila Basto Fagundes

Universidade Federal de Santa Catarina, UFSC, Email: priscila.fagundes@posgrad.ufsc.br

Douglas Dyllon Jeronimo de Macedo

Universidade Federal de Santa Catarina, UFSC, Email: douglas.macedo@ufsc.br

Moisés Lima Dutra

Universidade Federal de Santa Catarina, UFSC, Email: moises.dutra@ufsc.br

RESUMO

A gestão da qualidade sempre foi considerada uma atividade crítica durante o processo de gerenciamento dos dados e da informação. Com o advento dos ambientes big data, caracterizados pelos grandes volumes de dados e vasta variedade e complexidade, a forma como os dados eram gerenciados e transformados em informação precisaram sofrer adaptações de maneira a assegurar a qualidade, melhorar a disponibilidade e suas formas de representação. Este artigo tem como objetivo apresentar uma análise das relações existentes entre os critérios de avaliação da qualidade da informação referentes a acessibilidade, suficiência, credibilidade, completeza, representação concisa, representação consistente, facilidade de operação, exatidão, interpretabilidade, objetividade, relevância, reputação, segurança, atualidade, compreensibilidade e os aspectos que caracterizam os sistemas que envolvem o conceito de big data, identificados como volume, variedade, velocidade, veracidade, valor, variabilidade e visualização. Esta análise pretende fornecer um embasamento teórico acerca dos temas e colaborar com pesquisas que visam a proposição de modelos para avaliação da qualidade da informação específicos para ambientes big data. A partir da análise realizada, foi possível verificar a existência de relações entre os critérios da qualidade da informação que fizeram parte do estudo e os aspectos do big data, bem como a necessidade de que os modelos propostos contemplem as características desses ambientes para serem considerados eficientes. Esta pesquisa é identificada como bibliográfica de caráter exploratório, uma vez que foi realizado um estudo acerca dos conceitos relacionados a big data e qualidade da informação, e quanto a sua natureza é considerada uma pesquisa básica, pois objetiva a geração de novos conhecimentos que sejam úteis para o avanço da ciência, sem aplicação prática prevista.

Palavras-chave: Qualidade da Informação. Big Data. Critérios de Qualidade da Informação. V's do Big Data.

1 INTRODUÇÃO

A informação desde o final da Segunda Guerra Mundial, vem sendo percebida cada vez mais como um recurso importante e indispensável para as organizações. Os problemas relacionados com a sua localização e armazenamento, garantindo que ela circule de maneira a atender todas as necessidades dos seus usuários é visto como um grande desafio a ser considerado. De acordo com Araújo (2014), as primeiras reflexões sobre a gestão da informação incidiram sobre a sua natureza

física que consistia em reduzir o excesso, otimizar a circulação, identificar com precisão quais eram indispensáveis e descartar as irrelevantes ou duplicadas. Porém, o tratamento e uso da informação pela sociedade, vêm sofrendo modificações principalmente devido ao surgimento de novos modelos sociais, econômicos e tecnológicos (RIBEIRO, 2014).

No que diz respeito a qualidade da informação, esta é considerada um dos pilares fundamentais para a sobrevivência das organizações em um mercado cada vez mais competitivo. A informação passou a ser percebida como um recurso, uma condição de produtividade para as empresas. Questões como avaliar, mensurar e melhorar a qualidade da informação, tem sido o tema de pesquisa de vários autores, com as mais diferentes abordagens e propostas (CALAZANS, 2008). Com o crescimento dos dados gerados pelas organizações e a necessidade de informações de qualidade, tornou-se inevitável a criação de novas soluções e tecnologias que auxiliassem na gestão dos mesmos. Neste contexto, surge o conceito de *big data*, propondo novas abordagens e procedimentos para a geração, seleção e manipulação dos dados. Campos técnicos e do conhecimento foram afetados por esse novo ambiente, o que gerou o desenvolvimento de estudos e aplicações em diferentes áreas científicas, como a computação, as engenharias, as ciências sociais, entre outras (FURLAM; LAURINDO, 2017).

Com o objetivo de contribuir com novos estudos na área da qualidade da informação em ambientes *big data*, este artigo se propõe a realizar uma análise das relações entre os critérios da qualidade da informação, propostos por Lee *et al.* (2002): acessibilidade, suficiência, credibilidade, completeza, representação concisa, representação consistente, facilidade de operação, exatidão, interpretabilidade, objetividade, relevância, reputação, segurança, atualidade e compreensibilidade, e as características associadas aos ambientes *big data* conhecidas na literatura como os sete V's, volume, variedade, velocidade, veracidade, valor, variabilidade e visualização.

A partir dos resultados obtidos sobre pesquisas relacionadas ao tema qualidade da informação em *big data*, e que serão apresentados na sessão 3 deste artigo, chegou-se a conclusão que estudos que fazem referência ao tema são incipientes e ao mesmo tempo necessários no contexto atual. Desta forma, pretende-se auxiliar pesquisadores da área da qualidade da informação no entendimento das relações entre os critérios da qualidade da informação e os aspectos que envolvem os ambientes *big data*, e contribuir para que futuramente sejam propostos modelos para avaliação da qualidade da informação aplicáveis especificamente neste tipo de ambiente.

2 QUALIDADE DA INFORMAÇÃO E BIG DATA

2.1 QUALIDADE DA INFORMAÇÃO

As discussões acerca da qualidade da informação tiveram início em 1989, em um seminário ocorrido em Copenhaga na Dinamarca e promovido pelo Nordic Council for Scientific Information and Research Libraries (NORDINFO). A publicação decorrente do encontro tornou-se uma importante referência em relação ao assunto e foi reconhecida como o esforço mais importante de teorização sobre o tema na época (PAIN; NEHMY; GUIMARÃES, 1996).

Nehmy e Pain (1998), consideram a qualidade da informação de duas formas, uma delas tende a pensar na avaliação da qualidade a partir do sistema (do conteúdo, da informação, do intermediário ou da instituição), considerando aspectos objetivos, e a outra procura focar no usuário, enfatizando aspectos subjetivos da busca pela informação. Porém, é percebido que embora exista o anseio de se atribuir caráter objetivo aos aspectos relacionados ao sistema, ao final, a qualidade só teria sentido no contexto de sua utilidade para o usuário. Sendo assim, conforme as autoras, a questão da qualidade da informação - entendida enquanto excelência - estaria determinada no momento da interação do usuário com o sistema ou produto.

De acordo com Valente e Fujino (2016), embora ainda não haja um consenso sobre o conceito de qualidade da informação, observa-se na Ciência da Informação diversas proposições de critérios ou atributos para qualificar a informação. Tais critérios são usados para definir, medir e gerenciar a qualidade dos dados e informações e variam de acordo com as abordagens e vertentes sob as quais os estudos são realizados e de acordo com o contexto onde o conceito de qualidade é aplicado e avaliado. A variedade de atributos propostos para a qualidade da informação pode ser observada em um estudo apresentado em Arouck (2011). O pesquisador realizou um levantamento das publicações feitas entre 1974 e 2009 sobre o tema, onde identificou a proposição de 40 atributos da qualidade da informação distintos e propôs um agrupamento dos mesmos de acordo com três categorias: categoria meio, que corresponde ao nível técnico da informação; categoria de conteúdo, que diz respeito ao nível semântico da informação e a categoria de uso, que corresponde ao nível de eficácia ou influência da informação. Porém, vale ressaltar que a pesquisa de Arouck (2011), apenas categoriza os atributos, não sugerindo o contexto onde os mesmos devem ser aplicados.

A partir da década de 90, foram realizadas diversas proposições de modelos contendo diretrizes para a aplicação de critérios na qualidade da informação. Uma delas foi feita por um grupo de pesquisadores americanos, incluindo Richard Y. Wang, ligado ao grupo Massachusetts Institute of Technology of Information Quality (MITIQ), denominada Methodology for Information Quality Assessment (AIMQ), que apresenta um modelo para avaliação da qualidade da informação formado por três componentes (LEE *et al.*, 2002). O primeiro componente é formado pelos critérios de qualidade que farão parte do processo de avaliação. O segundo componente é um conjunto de diretrizes em forma de questionário para medir a qualidade da informação de acordo com os critérios que serão avaliados. E por fim, o terceiro componente fornece subsídios para a análise e interpretação dos resultados da avaliação. Os critérios de avaliação que fazem parte do AIMQ foram baseados em outros modelos propostos anteriormente e definidos a partir das necessidades observadas pelos pesquisadores dentro de organizações que possuem a informação como objeto do seu negócio. A seguir serão apresentados os critérios para qualidade da informação que fazem parte de AIMQ, de acordo com Lee *et al.* (2002).

- J Acessibilidade: indica em que medida as informações estão disponíveis e podem ser acessadas;
- J Suficiência: indica se a informação é suficiente para as necessidades dos usuários;
- J Credibilidade: remete ao grau de veracidade de um determinado conteúdo informacional;
- J Completeza: relaciona-se ao conjunto de informações para representar uma determinada realidade com todos os elementos descritivos necessários;
- J Representação Concisa: indica se a informação é apresentada de forma compacta e objetiva;
- J Representação Consistente: se refere à estruturação formal da informação, remetendo a utilização de padrões para a sua representação;
- J Facilidade de operação: indica se a informação é de fácil manipulação e combinação com outras informações;
- J Exatidão: está relacionada ao grau de correção da informação;
- J Interpretabilidade: indica se a informação possui clareza e é de fácil interpretação;
- J Objetividade: diz respeito a não contaminação de uma determinada informação por visões ou interesses particulares, referindo-se à imparcialidade;

- J Relevância: indica o grau de aplicabilidade e utilidade da informação;
- J Reputação: se refere a confiabilidade da informação e se a mesma provém de fontes confiáveis;
- J Segurança: indica o grau de segurança em relação a acessibilidade da informação, se ela é protegida de acessos indevidos;
- J Atualidade: se refere à novidade e atualização das informações disponibilizadas;
- J Compreensibilidade: indica se a informação é de fácil entendimento e compreensão.

O conjunto de critérios acima foi elencado para fazer parte da análise que será apresentada neste artigo, e a justificativa pela sua escolha, dentre os outros existentes na literatura, se deu por serem atributos utilizados na avaliação da qualidade da informação a partir da visão dos consumidores da informação, e também pela relevância da publicação que propõe o modelo. O artigo sobre o AIMQ (LEE *et al.*, 2002) foi publicado no periódico “*Information & Management*”, considerado um importante periódico na área da Ciência da Informação e que até o momento da escrita deste artigo possuía, segundo a ferramenta Google Scholar, mais de 1.330 citações.

A qualidade da informação pode ser considerada como uma fundamental e crescente área de estudos, envolvendo não somente os profissionais da informação, mas também profissionais advindos de outras áreas do conhecimento. Devido ao volume exponencialmente crescente de dados veiculados, principalmente na internet e o surgimento de novas tecnologias para atender cenários com particularidades específicas, os critérios para qualidade da informação também devem ser considerados e aplicados de acordo com as novas necessidades. E para que o processo de gestão da qualidade e avaliação da informação seja eficiente e resulte em informações de qualidade, é fundamental que novos modelos de qualidade da informação sejam propostos (LOPES, 2004).

2.2 BIG DATA

De acordo com McAfee e Brynjolfsson (2012), o termo *big data* está relacionado a grandes volumes de dados complexos, que possuem características distintas, são heterogêneos e que se originam de diferentes fontes, com controles distribuídos e descentralizados. Para Erl, Khattak e Buhler (2016), *big data* é um campo que se dedica à análise, ao processamento e ao

armazenamento de grandes *datasets*¹, e que as soluções e práticas de *big data* são geralmente necessárias quando as tecnologias e técnicas tradicionais não são suficientes para a execução destas atividades.

Cada vez mais as organizações estão gerando uma enorme quantidade de dados que são recebidos a partir de várias fontes e armazenados de diferentes maneiras, o que demanda um processo de gestão específico para garantir a sua qualidade. Vianna, Dutra e Frazzon (2016), enfatizam a importância de se fazer uma gestão efetiva e a [...] “necessidade de transformar esses dados em informações de qualidade, que possam ser utilizadas para direcionar os negócios e as estratégias das organizações, minimizando riscos, e apoiando o processo de tomada de decisões”. Para Erl, Khattak e Buhler (2016), *big data* não é apenas tecnologia é também sobre como as tecnologias podem impulsionar uma organização.

Existem na literatura, diferentes pontos de vista em relação as características que compõem um ambiente *big data*, o que parece consenso é que para que um conjunto de dados seja considerado *big data* ele deve possuir pelos menos uma dessas particularidades. Três delas foram inicialmente identificadas por Doug Laney no início de 2001, quando o autor publicou um artigo descrevendo o impacto do volume, velocidade e variedade de dados em *data warehouses* corporativos. Anos mais tarde novas características foram incorporadas ao conjunto de aspectos relacionados a *big data*, são elas: veracidade, valor, variabilidade e visualização (ZIKOPOULOS *et al.*, 2012; GANDOMI; HAIDER, 2015).

Para fins da análise em relação aos critérios da qualidade da informação, objetivo deste artigo, serão levados em consideração os sete V's citados anteriormente, sendo eles: Volume, Variedade, Velocidade, Veracidade, Valor, Variabilidade e Visualização. Buscando auxiliar em um melhor entendimento acerca dos aspectos que serão analisados, a seguir serão apresentados um detalhamento dos mesmos.

A primeira característica é o Volume, e refere-se a grande quantidade de dados e informações que são gerados a partir de fontes variadas. De acordo com Erl, Khattak e Buhler (2016), o volume de dados que é processado pelas soluções *big data* é substancial e cada vez maior, e impõem exigências distintas de armazenamento e processamento. Esta característica apresenta um

¹ *Datasets* são coleções ou agrupamentos de dados relacionados, onde cada grupo ou membro do grupo compartilha as mesmas propriedades ou atributos (ERL; KHATTAK; BUHLER, 2016).

grande desafio às estruturas de TI convencionais, pois grandes volumes de dados requerem armazenamento escalonável e uma abordagem distribuída para possibilitar a consulta aos mesmos.

A característica relacionada a Variedade diz respeito a diversidade dos dados e informações. É possível encontrar dados em diferentes lugares e formatos e essa variedade precisa ser gerida como um todo, uma vez que é possível que um determinado dado seja inútil se não for associado a outro, trazendo desafios para as organizações em termos de integração, transformação, processamento e armazenamento. Kaisler *et al.* (2013), afirma que a partir de uma perspectiva analítica, fazer a gestão da variedade de dados é provavelmente o maior obstáculo para utilização efetiva de grandes volumes de dados. Formatos de dados incompatíveis, estruturas de dados não alinhadas e semânticas inconsistentes representam desafios significativos que se não vencidos podem levar ao insucesso de um projeto.

O terceiro “V”, se refere a Velocidade, e possui relação com o tempo de resposta para determinada requisição. Com as comunicações em tempo real, cada vez mais, vem sendo possível atingir uma maior velocidade para troca de dados e informação (RIBEIRO, 2014). McAfee e Brynjolfsson (2012) afirmam que para muitas aplicações, a velocidade de criação de dados é ainda mais importante do que o volume dos mesmos, uma vez que as informações em tempo real ou quase em tempo real, tornam possível que uma empresa seja mais ágil que seus concorrentes.

No que diz respeito a quarta característica, a Veracidade está relacionada à qualidade ou à confiabilidade dos dados. Uma definição acerca deste aspecto foi apresentada por Erl, Khattak e Buhler (2016), que afirmam que os dados precisam ser avaliados quanto à confiabilidade, o que pode demandar atividades específicas para identificar os que não atendem a essa premissa e removê-los dos *datasets*. Um dado considerado de má qualidade é aquele que não pode ser convertido em informação e, portanto, não tem valor, enquanto que dados de qualidade possuem valor e geram informações significativas. Os autores complementam que os dados que são adquiridos de forma controlada, por exemplo através de registros de clientes on-line, geralmente contém mais qualidade do que os dados adquiridos por meio de fontes não controladas, como postagens em blogs ou redes sociais. Sendo assim, pode-se dizer que a qualidade dos dados está relacionada com a fonte e o tipo dos mesmos.

O “V” de Valor está relacionado com o retorno do investimento e é o resultado da combinação dos aspectos citados anteriormente. Esta característica está intuitivamente relacionada com a característica de veracidade, pois quanto maior a qualidade dos dados, mais valor ela possui

para o negócio (KAISLER *et al.*, 2013). Da mesma forma, pode-se considerar que valor e velocidade possuem relações inversas, pois quanto mais tempo se leva para que os dados sejam transformados em informações relevantes, menos valor terá para o negócio, uma vez que resultados obsoletos prejudicam a qualidade e a rapidez na tomada de decisão.

Os dois últimos V's dizem respeito a Variabilidade e a Visualização. O primeiro refere-se aos dados que estão em constante variação, como por exemplo, dados meteorológicos. Gandomi e Haider (2015), afirma que a variabilidade possui relação com a variação nos fluxos de dados. Já a Visualização consiste na preocupação com a utilização de técnicas para a representação das informações. Uma vez que o dado é processado, é necessário apresentar a informação gerada de forma clara e objetiva.

Com as características apresentadas é possível determinar a noção do termo *big data*. Porém, conforme mencionado anteriormente, especialistas acreditam que não há a necessidade da existência de todos os fatores ao mesmo tempo para que a solução seja considerada como tal, visto que existem casos em que há um maior destaque para uma ou outra característica, e em outros casos elas não são identificadas.

3 PROCEDIMENTOS METODOLÓGICOS

Quanto aos procedimentos metodológicos utilizados na escrita deste artigo, pode-se dizer que o mesmo trata de um estudo bibliográfico de caráter exploratório, pois pretende investigar os conceitos relacionados a *big data*, bem como a qualidade da informação levando em consideração seus aspectos e critérios respectivamente (GIL, 2007). Segundo Sampieri (1994), “o uso do estudo exploratório se dá quando o objetivo é examinar um determinado tema que tenha sido pouco ou nada estudado anteriormente, permitindo que se obtenha um maior grau e familiaridade com os fenômenos envolvidos” (apud VIANNA; DUTRA; FRAZZON, 2016). O estudo foi realizado a partir do levantamento e leitura de referências teóricas, publicadas por meios escritos e eletrônicos como livros, artigos científicos e páginas de web sites. Quanto a natureza, esta pesquisa é considerada uma pesquisa básica, pois objetiva a geração de novos conhecimentos que sejam úteis para o avanço da ciência, sem aplicação prática prevista.

A fim de verificar a relevância da pesquisa e a sua possível contribuição para o meio acadêmico, no dia 21/04/2017, foi realizada uma busca pelos termos “*big data*” e “*information quality*” utilizando o campo de pesquisa Tópico, que engloba a consulta ao título, ao resumo e as

palavras-chave, nos artigos publicados em periódicos nas bases de dados Scopus, Web of Science, Library & Information Science Abstracts (LISA) e Brapci. Após a busca, foi feita a leitura dos resumos destes artigos e selecionados apenas os que tratavam especificamente dos dois temas, totalizando 11 publicações, porém nenhuma delas abordando os temas qualidade da informação e *big data* no contexto dos seus critérios e aspectos respectivamente. A base de dados Scopus, retornou 6 artigos para os termos, no entanto apenas 1 deles, escrito por Tatiana Lukoianova e Victoria L. Rubin (2014), tratava especificamente da relação entre as dimensões da qualidade da informação e as características de ambientes *big data*, porém contemplando apenas o critério veracidade. Os mesmos termos foram utilizados para busca na base de dados Web of Science, a qual retornou 5 resultados, sendo que nenhum deles abordando os temas foco deste artigo. Nas bases de dados LISA e Brapci não foi encontrado nenhum trabalho contendo os termos “*big data*” e “*information quality* /qualidade da informação”.

Vale ressaltar que, por concordar com a definição de Somasundaram e Shrivastava (2011), na qual os dados são um conjunto de valores ou ocorrências, que após processados geram as informações, optou-se, para este estudo a não utilização do termo “*data quality*” ou qualidade de dados, como opção de busca nas bases de dados citadas anteriormente.

4 OS CRITÉRIOS DA QUALIDADE DA INFORMAÇÃO E OS V's DO BIG DATA

A seguir serão apresentados os relacionamentos identificados entre os critérios da qualidade da informação e os aspectos que caracterizam os ambientes *big data* e logo após uma análise dos mesmos. Os relacionamentos foram estabelecidos a partir dos critérios que fazem parte da metodologia AIMQ proposta em Lee et al. (2002) pelos motivos já apresentados anteriormente. Em relação as características *big data*, os aspectos selecionados foram: volume, velocidade, variedade, valor, veracidade, variabilidade e visualização e a escolha pela utilização dos sete V's, se deu por se acreditar que este conjunto abrange de forma ampla as particularidades que envolvem este conceito.

Procurou-se estabelecer uma relação entre os critérios da qualidade da informação e os aspectos *big data* que apresentassem objetivos comuns e que de alguma forma fossem complementares. O Quadro 1, mostra um resumo do resultado das relações identificadas e logo após serão apresentadas as justificativas para cada uma delas.

Quadro 1- Relação entre os critérios da qualidade da informação e os V's do *big data*

CRITÉRIOS DA QUALIDADE DA INFORMAÇÃO \ V'S DO BIG DATA	VOLUME	VELOCIDADE	VARIEDADE	VALOR	VERACIDADE	VARIABILIDADE	VISUALIZAÇÃO
ACESSIBILIDADE		X		X			X
SUFICIÊNCIA	X		X	X			
CREDIBILIDADE				X	X	X	
COMPLETEZA	X		X	X			
REPRESENTAÇÃO CONCISA				X			X
REPRESENTAÇÃO CONSISTENTE		X	X	X			X
FACILIDADE DE OPERAÇÃO			X	X			
EXATIDÃO			X	X	X	X	
INTERPRETABILIDADE				X			X
OBJETIVIDADE				X			
RELEVÂNCIA				X	X		
REPUTAÇÃO			X	X	X		
SEGURANÇA				X			X
ATUALIDADE	X	X		X			
COMPREENSIBILIDADE				X			X

Fonte: Elaborado pelos autores.

Acredita-se, que o aspecto volume do *big data*, possui relacionamento com os critérios suficiência, completeza e atualidade. Os relacionamentos entre o volume e os critérios suficiência e completeza possuem justificativas semelhantes, uma vez que entende-se que para que as informações sejam consideradas suficientes é necessário que se tenha quantidades de dados disponíveis. Apesar também de entender que grandes volumes de dados não garantam informações suficientes para se alcançar um determinado objetivo, acredita-se que a probabilidade de se encontrar a informação desejada e que a mesma seja eficaz em um universo de dados maior, seja

mais alta que, se o mesmo universo não possuir dados suficientes disponíveis. Já a relação entre volume dos dados e informações atualizadas foi estabelecida, por acreditar que informações atualizadas constantemente geram grandes quantidades de dados.

Quanto ao aspecto velocidade, este possui relação com três critérios da qualidade da informação. São eles: atualidade, pois para as informações estarem atualizadas e serem consideradas atuais, uma das necessidades é que o tempo de resposta para as requisições seja pequeno, informações consideradas atuais possuem ligação com o intervalo de tempo em que os dados são processados e as informações disponibilizadas; acessibilidade, pois garantir o acesso as informações de forma rápida e eficiente exige tecnologia que possibilite agilidade neste processo e ambientes *big data* devem considerar o uso de tecnologias que permitam velocidade tanto no processamento dos dados como na disponibilização das informações e por último o critério representação consistente, uma vez que acredita-se que o formato utilizado para a informação apresentada para o usuário, pode influenciar no tempo em que o dado levará para ser estruturado e transformado em informação.

A variedade, como uma característica do *big data*, está relacionada com os critérios completude e suficiência pois entende-se que informações consideradas completas significam que apresentam integridade e sugerem a inclusão de todos os dados necessários relativo a um determinado problema. Está relacionada também com o critério representação consistente, uma vez que ambientes *big data* se caracterizam por possuírem uma grande variedade de dados que podem possuir diferentes formatos, e garantir a integridade das informações a partir dessa variedade pode ser um fator desafiador e se não considerado pode comprometer a qualidade da informação. Da mesma forma, o critério que foca na facilidade de operação das informações, pois a variedade dos dados pode interferir na combinação das informações e manipulação das mesmas.

Ainda a respeito dos critérios que se relacionam com o aspecto variedade do *big data*, tem-se a exatidão e a reputação. Dados que apresentam uma grande variedade e são providos por fontes distintas demandam um cuidado especial para garantir que as informações geradas e disponibilizadas sejam exatas e confiáveis.

Em relação a característica valor, a análise propõe que todos os critérios da qualidade da informação estão relacionados com esse aspecto dos ambientes *big data*. Conforme apresentado anteriormente, esta característica está ligada ao retorno do investimento que é feito pela organização na implementação de ambientes *big data*. Da mesma forma, se todos os critérios da

qualidade da informação forem atendidos, a probabilidade de o investimento gerar valor para a organização é bem maior que o contrário. Sistemas que geram informações, íntegras e suficientes (completeza e suficiência), disponíveis (acessibilidade), seguras (segurança), atuais (atualidade), corretas e confiáveis (exatidão, credibilidade e reputação), úteis (relevância), imparciais (objetividade), que sejam de fácil visualização e manipulação (representação concisa, representação consistente e facilidade de operação), e de fácil compreensão (interpretabilidade e compreensibilidade), serão considerados de qualidade para os seus usuários e conseqüentemente darão retorno para o investimento feito pela organização.

Os critérios da qualidade da informação, exatidão e credibilidade possuem descrições bastante semelhantes e estão relacionados com o aspecto veracidade do *big data* pois ambos se referem ao grau de confiabilidade da informação gerada a partir dos dados armazenados, uma vez que dados não confiáveis não geram informações corretas. Outra relação com esta característica é o critério de relevância, uma vez que este, indica o grau de aplicabilidade da informação em um determinado contexto, conteúdo informacionais não confiáveis, não são uteis, comprometendo assim a sua aplicabilidade e utilização. E por fim a relação entre a veracidade e a reputação, pois parte-se do princípio que fontes seguras geram informações confiáveis.

No que diz respeito a variabilidade, acredita-se que da mesma forma que o aspecto citado anteriormente, seu relacionamento está com os critérios, exatidão que diz respeito a disponibilização de informações corretas, bem como a credibilidade que remete a confiabilidade do conteúdo informacional. Informações que possuem um grande grau de variabilidade, necessitam de um tratamento adequado para garantir exatidão e credibilidade e conseqüentemente a qualidade.

E para finalizar, acredita-se que o aspecto visualização, possui relação com os critérios acessibilidade e segurança, pois as informações precisam estar disponíveis, mas ao mesmo tempo protegidas contra acessos indevidos. Com os critérios que dizem respeito a representação concisa e consistente, uma vez que tratam sobre a forma em que a informação deve ser apresentada ao usuário. E também com os critérios interpretabilidade e compreensibilidade que indicam que a informação deve ser apresentada de maneira clara, permitindo uma fácil interpretação e entendimento.

Como pode ser observado foi possível identificar relações entre todos os critérios de qualidade da informação elencados a partir da metodologia AIMQ e os sete V's que contemplam os aspectos *big data*. Porém, ainda não é possível afirmar com base nas relações criadas se é possível a

proposição de um modelo de qualidade da informação eficiente apenas com os critérios de qualidade que foram utilizados nesta análise. A seguir serão apresentadas as considerações finais sobre este estudo, bem como proposições de trabalhos futuros que possam contribuir para a estruturação de modelos de qualidade específicos para serem aplicados em contextos *big data*.

5 CONSIDERAÇÕES FINAIS

Conforme apresentado na introdução deste artigo, pesquisas relacionadas à qualidade da informação em ambientes *big data* são incipientes na literatura em geral e, particularmente na área da Ciência da Informação e apresentam-se como uma oportunidade de estudo. Sendo assim, considera-se que trabalhos que contemplem os dois temas, necessitam de novas pesquisas sobre novas técnicas na área da gestão da informação em ambientes *big data*, tratando especificamente da qualidade da informação.

Como sugestões de trabalhos futuros, propõem-se que seja realizada uma análise de quais outros atributos da qualidade da informação podem ser relacionados com os aspectos *big data*, pois acredita-se que este estudo será de fundamental importância para a proposição de um modelo de qualidade da informação específico para *big data*. Vale considerar a possibilidade de que, não necessariamente deverá ser proposto um único modelo, mas um para cada aspecto, uma vez que, conforme visto anteriormente não é necessário que sejam identificados os sete V's para um ambiente ser considerado *big data*. Desta forma será possível avaliar a qualidade da informação em sistemas *big data* de acordo com as características que ele apresenta.

An Analysis between the Information Quality and Big Data

ABSTRACT

Quality management has always been considered a critical activity during the data and information management process. With the advent of big data environments, characterized by large volumes of data, vast variety and complexity, the way data was managed and transformed into information needed to be adapted in order to ensure quality, improve availability and its forms of representation. This article aims to present an analysis of the relationship between the criteria for assessing information quality: accessibility, appropriate amount, believability, completeness, concise representation, consistent representation, ease of operation, free-of-error, interpretability, objectivity, relevancy, reputation, security, timeliness, understandability and the aspects that characterize the systems that involve the concept of Big Data: volume, variety, velocity, veracity, value, variability and visualization. This analysis intends to provide a theoretical basis on the themes and to assist other researchers in the proposition of information quality assessment models specific to Big Data. From the analysis, it was possible to conclude that there is a

relationship between all the information quality criteria that were part of the study and the big data aspects, and it is fundamental that the models proposed for this context contemplate the characteristics of these environments to be considered efficient. This research is considered an exploratory bibliography, once a study was carried out about the concepts related to the big data and information quality, it is considered a basic research, since it aims to generate new knowledge that are useful for the advancement of science, without practical application envisaged.

Keywords: Information quality. Big Data. Criteria of information quality. V's of Big Data.

REFERÊNCIAS

ARAÚJO, Carlos Alberto Ávila. Fundamentos da ciência da informação: correntes teóricas e o conceito de informação. **Perspectivas em Gestão & Conhecimento**, v. 4, n. 1, 2014, p. 57-79. Disponível em: <<http://periodicos.ufpb.br/ojs2/index.php/pgc>>. Acesso em: 10 mar. 2017.

AROUCK, Osmar. Atributos de qualidade da informação. In: XII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 12., 2011, Brasília. **Anais eletrônicos...** Brasília, 2011. Disponível em: <<http://enancib.ibict.br/index.php/enancib/enancibXII/paper/view/874>>. Acesso em: 25 maio 2017.

CALAZANS, Angélica Toffano Seidel. Qualidade da informação: conceitos e aplicações. **Transinformação**, v. 20, n. 1, 2008, p. 29-45. Disponível em: <<http://www.scielo.br/pdf/tinf/v20n1/03.pdf>>. Acesso em: 3 mar. 2017.

ERL, Thomas; KHATTAK, Wajid; BUHLER, Paul. **Big Data fundamentals: Concepts, Drivers & Techniques**. Boston: Prentice Hall, 2016.

FURLAN, Patricia Kuzmenko; LAURINDO, Fernando José Barbin. Agrupamentos epistemológicos de artigos publicados sobre big data analytics. **Transinformação**, v. 29, n. 1, 2017, p. 91-100. Disponível em: <<http://www.scielo.br/pdf/tinf/v29n1/0103-3786-tinf-29-01-00091.pdf>>. Acesso em: 21 abr. 2017.

GANDOMI, Amir; HAIDER, Murtaza. Beyond the hype: Big data concepts, methods, and analytics. **International Journal of Information Management**, v. 35, n. 2, 2015, p. 137-144. Disponível em: <<http://dx.doi.org/10.1016/j.ijinfomgt.2014.10.007>>. Acesso em 21: abr. 2017.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

KAISLER, Stephen *et al.* Big Data: Issues and Challenges Moving Forward. In: XLVI HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 46., Maui, 2013. **Anais eletrônicos...** Maui, 2013. Disponível em: <<https://www.computer.org/csdl/proceedings/hicss/2013/4892/00/4892a995.pdf>>. Acesso em: 22 abr. 2017.

LEE, Yang W. *et al.* AIMQ: A methodology for information quality assessment. **Information and Management**, v. 40, n. 2, 2002, p. 133-146. Disponível em:

<<http://mitiq.mit.edu/Documents/Publications/TDQMpub/2002/AIMQ.pdf>>. Acesso em: 20 maio 2017.

LOPES, Ilza Leite. Novos paradigmas para avaliação da qualidade da informação em saúde recuperada na Web. **Ciência da Informação**, v. 33, n. 1, 2004, p. 81-90. Disponível em: <<http://revista.ibict.br/ciinf/article/view/1071/1165>>. Acesso em: 13 maio 2017.

LUKOIANOVA, Tatiana; RUBIN, Victoria L. Veracity roadmap: Is big data objective, truthful and credible? **Advances in Classification Research Online**, v. 24, 2014, p. 4-15. Disponível em: <<http://journals.lib.washington.edu/index.php/acro/article/view/14671/12311>>. Acesso em: 21 abr. 2017.

MCAFEE, Andrew; BRYNJOLFSSON, Erik. Big Data. The management revolution. **Harvard Business Review**, v. 90, n. 10, 2012 p. 61–68. Disponível em: <<https://hbr.org/2012/10/big-data-the-management-revolution>>. Acesso em: 22 abr. 2017.

NEHMY, Rosa Maria Quadros; PAIM, Isis. A desconstrução do conceito de “qualidade da informação”. **Ciência da Informação**, v. 27, n. 1, 1998, p. 36–45. Disponível em: <<http://www.scielo.br/pdf/ci/v27n1/05.pdf>>. Acesso em: 28 mar. 2017.

PAIM, Isis *et al.* Problematização do conceito "Qualidade" da Informação. **Perspectivas em Ciência da Informação**, v. 1, n. 1, 1996, p. 111–119. Disponível em <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/8/27>>. Acesso em: 30 mar. 2017.

RIBEIRO, Claudio José Silva. Big Data: os novos desafios para o profissional da informação. **Informação & Tecnologia**, v. 1, n. 1, 2014, p. 96–105. Disponível em: <<http://periodicos.ufpb.br/index.php/itec/article/view/19380/11156>>. Acesso em: 19 abr. 2017.

SOMASUNDARAM G Siva.; SHRIVASTAVA Alok. **Armazenamento e Gerenciamento de Informações**: Como armazenar, gerenciar e proteger informações digitais. Porto Alegre: Bookman, 2011. 472p.

VALENTE, Nelma T. Zubek; FUJINO, Asa. Atributos e dimensões de qualidade da informação nas Ciências Contábeis e na Ciência da Informação: um estudo comparativo. **Perspectivas em Ciência da Informação**, v. 21, n. 2, 2016, p. 141–167. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2530/1761>>. Acesso em: 16 mar. 2017.

VIANNA, William Barbosa; DUTRA, Moisés Lima; FRAZZON, Enzo Morosini. Big data e gestão da informação: modelagem do contexto decisional apoiado pela sistemografia. **Informação & Informação**, v. 21, n. 1, 2016, p. 185. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/23327/18993>>. Acesso em: 21 abr. 2017.

ZIKOPOULOS, Paul. *et al.* **Understanding Big Data**: Analytics for Enterprise Class Hadoop and Streaming Data. New York: McGraw-Hill, 2012.