

Revisão sobre a qualidade de *Big Data* no contexto da Indústria 4.0

Paulo Henrique Amorim Santos (Universidade Federal do Paraná) phas223@hotmail.com

Izabel Cristina Zattar (Universidade Federal do Paraná) izabel.zattar@gmail.com

Resumo:

A necessidade de assegurar a qualidade de dados se intensifica ao gerenciar uma grande quantidade de informação. Existem obstáculos específicos para garantir a veracidade das informações coletadas, transmitidas e armazenadas em grande volume. Assim, com o desenvolvimento das tecnologias da Indústria 4.0 a pesquisa em *Big Data* evolui e se ramifica em novas áreas de estudo. Nesse contexto, o objetivo deste artigo é analisar criticamente a literatura relativa à análise da qualidade da tecnologia de *Big Data* na perspectiva da Indústria 4.0. A partir de uma abrangente revisão sistemática da literatura, é discutido a evolução do estudo perante as inovações recentes. Além disso, são identificados estudos de casos aonde as tecnologias já são utilizadas e a sua qualidade, avaliada. Dentre as necessidades identificadas, destacam-se a crescente preocupação em pesquisa e desenvolvimento de modelos estatísticos e computacionais, desenvolvimento de aprendizado de máquina e cuidados com segurança cibernética. A literatura também é convergente sobre a urgência em tornar acessível ao usuário a identificação de erros, através de processos transparentes. O estudo é direcionado à gestores, que buscam informações para projetos de implantação das tecnologias 4.0. Ainda, contribui-se com a comunidade acadêmica a partir da orientação à pesquisas futuras relativas ao tema.

Palavras chave: Qualidade de *Big Data*, Indústria 4.0, Análise de *Big Data*.

Review about *Big Data* Quality in the Industry 4.0 context

Abstract

The need to protect the big data quality intensifies when managing a lot of information. There are specific obstacles to ensure the accuracy of information collected, transmitted and stored in large volumes. With the development of Industry 4.0 technologies, Big Data research has evolved and branched out into new areas of study. In this context, the objective of this article is to critically analyze the literature corresponding to the quality analysis of Big Data technology from the perspective of Industry 4.0. From a comprehensive systematic review of the literature, we discuss the evolution of the study in the face of recent innovations. In addition, case studies are identified where technologies are already used and its quality is assessed. Among the identified needs, we highlight the growing concern in research and development of statistical and computational models, development of machine learning and cyber security care. The literature also converges on the urgency of making it accessible to the user to identify errors easily, through transparent processes. The study is aimed at managers, who seek information for projects implementing the 4.0 technologies. Furthermore, it contributes to the academic community by orienting future research on the subject.

Key-words: Big Data Quality, Industry 4.0, Big Data Analytics

1. Introdução

A Indústria 4.0 é um termo que foi usado pela primeira vez na Alemanha (originalmente Industrie 4.0), e é usado em todo o mundo para expressar os avanços tecnológicos que caracterizam a quarta revolução na fabricação industrial. (GRUSCHKA; LÜSSEM, 2016). Atualmente, as indústrias estão cada vez mais aderindo a avanços tecnológicos em cibernética, integração de recursos computacionais, infraestruturas de TI e novas tecnologias de obtenção e gerenciamento de dados. (REIS; KENETT, 2018). Essas novas tecnologias de monitoramento e controle do processo permitem atender as exigências atuais do mercado consumidor. (YAO; GE, 2018). Dentre as tecnologias mais difundidas citam-se a Internet das Coisas (IoT), sistemas ciber físicos (CPS), comunicação máquina-a-máquina (M2M), computação em nuvem e *Big Data*. (GRUSCHKA; LÜSSEM, 2016).

Neste ambiente, a qualidade da informação gerada é de importância central na Indústria 4.0. (REIS; KENETT, 2018). Para a indústria de processos, tanto a produtividade quanto a qualidade estão altamente relacionados à lucratividade e à segurança. (YAO; GE, 2018). Porém, assegurar a qualidade de dados é um grande desafio para qualquer organização, e é ainda mais desafiador no sentido de que as pessoas têm diferentes notações de definição e validação de dados. (MITTAL, 2013).

A má qualidade de *Big Data* está fadada a afetar os resultados de sua análise, aumentando o risco de falhas. (LAI, 2017). Assim, muitos sistemas de manufatura não estão prontos para gerenciar grandes volumes de dados devido às altas demandas em acesso a dados e qualidade. Como essas fontes de dados geralmente operam em ambiente heterogêneo, a integração entre os sistemas é problemático. (ALJUMAILI et al., 2015). Algoritmos preditivos terão dificuldades em distinguir sinal e ruído, se os dados forem bagunçados, e o processo de medição não estiver sob controle. (BENHENNI, 2017).

Assim, a qualidade de dados precisa ser assegurada para que os processos sejam confiáveis e significativos. Para garantir a qualidade e validade dos dados, é extremamente importante garantir que os dados corretos sejam usados. Isso requer padrões e procedimentos que tornam a informação compreensível, transferível e garante que a informação criada esteja disponível em todos os locais relevantes. (BERG; WILTS, 2018).

Neste contexto, este artigo objetiva analisar criticamente a literatura correspondente a análise de qualidade da tecnologia de *Big Data* na perspectiva da Indústria 4.0.

2. Fundamentação Teórica

Devido ao aumento contínuo de volume de dados, mais recursos são necessários para rápida execução confiável de tarefas de análise de dados. (HÖLSCHER et al., 2018). Enquanto as primeiras publicações de *Big Data* focavam em seu valor potencial, ultimamente o foco de sua pesquisa tem sido a aplicação de tecnologias e técnicas que buscam o valor comercial da utilização dos dados. (KAMDI, 2018).

Existem na literatura, diferentes pontos de vista em relação as características que compõem um ambiente de *Big Data*. Três delas foram inicialmente identificadas por Doug Laney no início de 2001: volume, velocidade e variedade dos dados (LANEY, 2001). Anos mais tarde novas características como veracidade, valor e variabilidade foram incluídas. (ZIKOPOULOS et al., 2012; GANDOMI; HAIDER, 2015).

A qualidade de *Big Data* enquadra-se na característica de veracidade, sendo relacionada à confiabilidade dos dados. (NGUYEN et al., 2018). 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. (FAGUNDES; MACEDO; FREUND, 2018).

A análise de *Big Data* compreende técnicas e tecnologias para capturar, armazenar, transferir, analisar e visualizar uma quantidade enorme de dados não estruturados. Esse processo está se tornando um ativo importante para os gerentes de negócios, devido à sua capacidade de criar informações e valor de conhecimento (VERMA; BHATTACHARYYA; KUMAR, 2018).

O gerenciamento de qualidade *Big Data* é crítico para os aplicativos de IoT, especialmente para a análise de dados. (GE, BANGUI, BUHNOVA, 2018). Assim, não restam dúvidas de que as aplicações da tecnologia de *Big Data* podem melhorar a vantagens competitivas de empresas e organizações. (LAI, 2017). *Big Data* assim, é um ativo importante para ajudar empresas a aumentar as vantagens da competitividade no mercado. (LAI, 2017)

Para Utz, Neumann e Tafreschi (2018), a qualidade dos dados é um dos principais requisitos para a tomada de decisões bem-sucedidas na Indústria 4.0. Trata-se de um fator importante no fornecimento de soluções de alta qualidade. Para Firmani et al. (2018), as dimensões para qualidade da informação podem ser agrupadas em clusters: precisão, completude, consistência, redundância, legibilidade, acessibilidade, confiança e utilidade. A qualidade dos dados é uma característica essencial que determina a confiabilidade dos dados para tomada de decisões. Novas tecnologias, como as bases de dados Hadoop, NoSQL e MPP, surgiram para resolver os desafios do *Big Data* e permitir que novos tipos de produtos e serviços sejam entregues pelas empresas. (ABDULLAH et al., 2015)

Neste contexto, o conteúdo da informação nos dados de inspeção deve ser alto. Para dados digitalizados, recursos limitados de armazenamento, manipulação e análise podem criar problemas em sua análise. Thoben, Wiesner e Wuest (2017) citam como desafios no gerenciamento de *Big Data* o aumento da complexidade, dinâmica do sistema, qualidade de dados, validação e verificação e a comunicação entre as máquinas. Testar o *Big Data* é um dos principais desafios que a indústria enfrenta hoje em dia. (KAMDI, 2018).

A literatura de *Big Data*, tanto acadêmica como profissional, tem forte foco nas oportunidades, mas pouca atenção às ameaças que resultam do aproveitamento errado de dados. (CLARKE, 2016). Além disso, existem requisitos para infraestruturas de *Big Data* que se diferenciam em cada setor. Problemas como exatidão, integridade, riqueza de contexto, disponibilidade e duração do arquivamento já foram identificados como áreas onde a indústria precisa melhorar a qualidade dos dados para análise avançada. (MOYNE; ISKANDAR, 2017). Os fabricantes precisam acessar e trabalhar facilmente com dados de fabricação, não importando onde estejam localizados ou que aplicativos os criam. (WOOL et al., 2018). Assim, há a necessidade de verificar a precisão dos dados, eliminando o ruído através de metodologias de limpeza de dados. (MAZAHUA et al. 2016). A limpeza de dados é o processo de detectar e remover entradas imprecisas em um conjunto de dados. (GALLAGHER et al., 2018). Isso é para garantir a qualidade dos dados para que as decisões que são feitas a partir dos dados coletados são precisas e eficazes. (MAZAHUA et al. 2016).

3. Metodologia

Para criar uma amostra seleta de publicações sobre o tema, primeiramente, foi definido o banco de dados para a busca de material e então selecionados os termos de busca.

O banco de dado escolhido foi o Google Acadêmico®. Todas as buscas foram limitadas à

resultados em inglês, português e espanhol, havendo limitação a artigos mais recentes que 2010. A definição dos termos de busca foi baseada no tema de estudo, buscando abranger uma grande gama de publicações. Os algoritmos de busca utilizados foram:

1. "data quality" AND "*Big Data*" AND ("industry 4.0" OR "factory 4.0" OR "smart manufacturing");
2. "*Big Data* quality"

Oriundos da base de dados foram considerados ao todo 1354 documentos para análise de alinhamento ao tema. Desta amostra inicial foram desconsiderados os documentos cujo texto completo não estava disponível em Open Source ou cujo idioma era incompatível com o critério adotado. Ainda foram excluídas as publicações cujo título e resumo não se mostraram relacionados ao tema da revisão. Como resultado foram obtidos 162 publicações que efetivamente relacionavam-se ao tema. A leitura completa das obras reduziu a amostra final à 66 publicações. A Figura 1 resume os passos adotados para a construção da amostra final de artigos considerados na revisão da literatura.

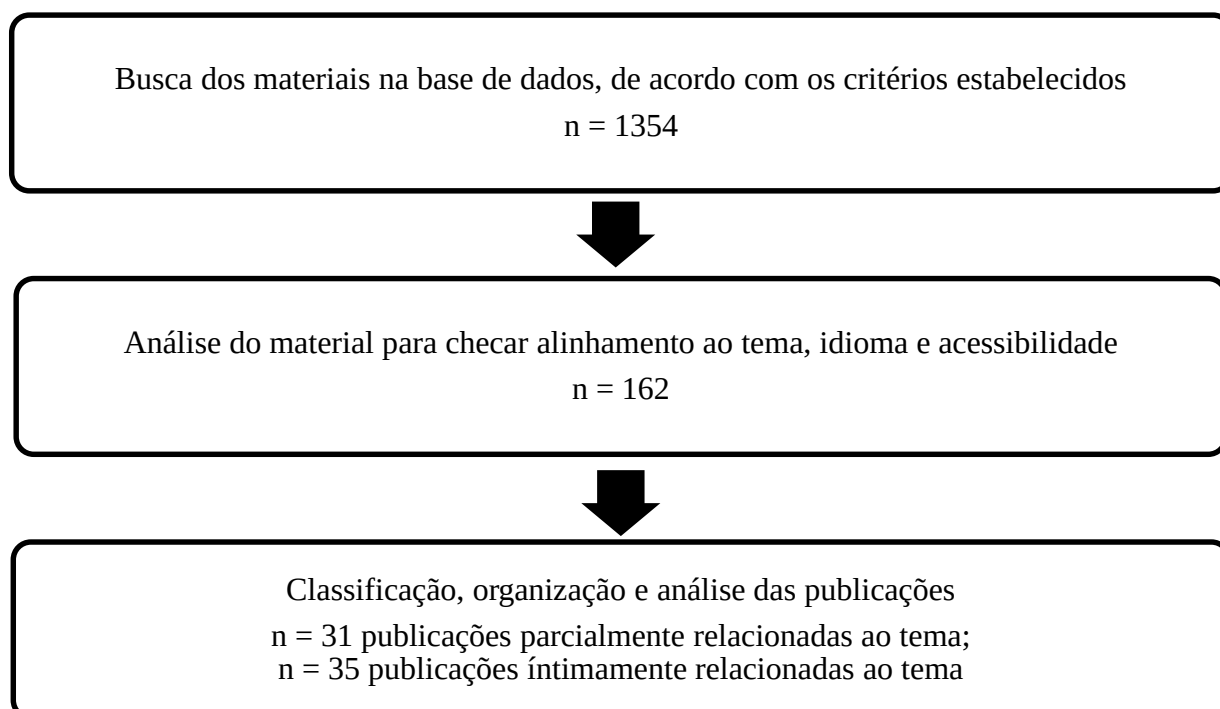


Figura 1 – Procedimento adotado para construção da amostra final de artigos estudados na revisão da literatura

A partir da amostra final, foi realizado uma análise crítica de cada publicação, e um estudo sobre o atual estado da análise de qualidade em *Big Data*. A Revisão Sistemática da Literatura apresentada neste artigo considera todas as publicações encontradas, tanto àquelas que são parcialmente relacionadas ao tema quanto trabalhos cujo foco é a investigação da qualidade de *Big Data*, de forma teórica ou prática.

4. Revisão Sistemática da Literatura

A pesquisa de qualidade de dados pode ser rastreada até os anos 80, mas com o advento da era de *Big Data*, a qualidade de *Big Data* tem vindo a emergir como uma nova área de pesquisa. (GE, BANGUI, BUHNOVA, 2018). Haug e Arlbjorn (2011) identificam um conjunto de

barreiras gerais para contra a garantia da alta qualidade de dados. Os resultados da pesquisa indicam que a falta de delegação de responsabilidades e a atualização dos dados são os aspectos que tem o maior impacto na qualidade dos dados. Além disso, a pesquisa mostra que a grande maioria das empresas acredita que sofre perdas devido a má qualidade dos dados.

Alguns dos trabalhos mais importantes da amostra datam de 2014: Hazen et al. (2014) introduziu o problema de qualidade dos dados no contexto da gestão da cadeia de abastecimento (SCM) e propôs métodos para monitoramento e controle da qualidade dos dados. No mesmo ano, Woodall et al. (2014) determinam empiricamente que a qualidade dos dados é fator da dimensão dos dados. Como resultado, os autores confirmam que, com base em um dos exemplos estudados, a qualidade dos dados tem um impacto cada vez mais negativo nos resultados de classificação, à medida que o tamanho do conjunto de dados aumenta. Contudo, o método de classificação, usado neste caso, produziu melhores resultados com um maior conjunto de dados de entrada para treinar o modelo, mesmo proporcionando maior quantidade de erros. Casale (2015) considera a questão de como a engenharia dirigida por modelos deve suportar os softwares de uso intensivo de dados sistemas. Os autores analisam alguns dos principais desafios envolvidos no desenvolvimento destas ferramentas e nos modelos subjacentes.

Li, Gong e Liu (2016) estudam como os métodos intensivos de dados de avaliação de ciclo de vida de materiais são dependentes da alta qualidade e confiabilidade dos dados. Utz, Neumann e Tafreschi (2018) apresentam em seu artigo um modelo de processo específico visando melhoria de disponibilidade de dados através da limpeza dos mesmo. Para eliminar o ruído e reconstruir valores faltantes, modelos de regressão e modelos probabilísticos assim como métodos de detecção discrepantes são aplicados. Madhikermi et al. (2016) abordam a Tomada de Decisão Multi-Critérios e a qualidade dos dados, dificilmente considerada nos sistemas de manutenção especializados existentes. No artigo, os autores desenvolvem um framework de avaliação de qualidade de relatórios de manutenção que permite facilmente avaliar qualidade de relatórios de manutenção em tempo real. O estudo de caso descreve como a ferramenta proposta está sendo usada por um fabricante de equipamentos multinacional finlandês para avaliar e aprimorar práticas de relatórios em um grupo específico ou em um grupo de filiais.

Khodabakhsh, Ari e Bakir (2017) focam em seu artigo no estudo da veracidade no setor de petróleo e gás. Os erros de medição nos dados do sensor utilizado são categorizados em duas classes principais pelos autores: erros aleatórios (ruído estatístico) e erros grosseiros (devido a problemas nos dispositivos sensores). Para os autores, os dados precisam ser limpos e reconciliados antes de poderem ser usados no treinamento de algoritmos de mineração. Já Bokrantz et al. (2017) contribuem no campo da simulação de eventos discretos, contribuindo com o estudo sobre empirismo e qualidade de dados na simulação de um processo fabril. Pawar e Thakore (2017) propõem um sistema que melhora a qualidade dos dados coletados para a tomada de decisões de negócios. O documento contribui para fornecer uma solução através da introdução de um novo modelo de avaliação para avaliar e gerenciar a qualidade dos dados de mídia sociais. Santos et al. (2017) propõem a implementação de *Big Data Analytics* em uma organização multinacional (Bosch Car Multimedia - Braga), como estudo de caso. Para garantir a qualidade dos dados, uma análise preliminar dos dados disponíveis foi feito, identificando os atributos que apresentam anomalias, como valores errôneos ou dados ausentes. Sadiq et al., 2017 delineam um chamado à ação para promover o empirismo em pesquisa de qualidade de dados. No estudo, os autores indicam a presença de um continuum de dados que varia de real para dados sintéticos, o que tem implicações sobre como os métodos de qualidade de dados são avaliados. Song et al., (2017) focam em dados defeituosos gerados pela operação inadequada

de componentes físicos e cibernéticos de um CPS de manufatura orientada a serviços (SMCPS), e desenvolvem políticas de gerenciamento efetivas para lidar com tais dados.

Reis e Kenett (2018) descrevem uma estrutura para avaliar, analisar e melhorar a qualidade das informações geradas em estudos empíricos, denominada InfoQ, utilizada no contexto da Indústria de Processamento Químico. Heinrich et al. (2018) propõem um conjunto de cinco requisitos para métricas de qualidade de dados para suportar a incerteza em uma gestão economicamente orientada a qualidade dos dados. Wahudi, Kuk e Janssen (2018) usam um estudo de caso de um grande empresa de telecomunicações para desenvolver um modelo padrão de processo genérico para melhorar a qualidade dos dados.

Dentro da amostra de artigos, ainda foram identificadas outras publicações com caráter de revisão de literatura sobre o tema. Abdullah et al. (2015) analisam as características da qualidade de *Big Data* e os processos de gerenciamento que estão envolvidos. Fagundes, Macedo e Freund (2018) apresentam um levantamento das publicações relacionadas à qualidade de dados em *Big Data* na base de dados Web of Science até o ano de 2016. Os autores apresentam um panorama geral das publicações relacionadas à qualidade de dados em ambientes de big, apresentando o local de publicação das pesquisas, uma síntese dos estudos, seus contextos de aplicação e principais autores. Biancone, Secinaro e Brescia (2018) definem qualidade de *Big Data* através de uma revisão e abordagem sistemática, entre janeiro de 2015 e dezembro 2016. O estudo investiga e fornece alguns pensamentos e ideias sobre o futuro da qualidade do *Big Data* e sua utilização na Governança em Saúde. O artigo ainda destaca a utilidade da qualidade do *Big Data* nas escolhas médicas em sistemas de informação.

A qualidade dos dados é crucial para obter o engajamento do usuário e a aceitação de serviços da Indústria 4.0. (DING et al., 2018). Porém, a maioria dos sistemas de automação não realiza verificações de qualidade nos dados e é difícil avaliar se os dados podem ser confiáveis ou não. (RØDSETH; PERERA; MO, 2016). Segundo Ding et al., a existência dos valores anormais e valores ausentes é principalmente devido a resposta incorreta ou não resposta.

Uma avaliação de disponibilidade de dados consiste em uma análise estatística inicial e de alto nível dos recursos do modelo proposto. Os resultados permitem que o profissional tome uma decisão informada, com base na qualidade e integridade dos dados. (GALLAGHER et al., 2018). Quando grande quantidade de dados está disponível para calcular o recurso identificado, poucos dados ausentes possuem um baixo impacto no recurso extraído. (LEE, KAO, YANG, 2014). Medir e garantir a qualidade de dados significa transparência do processo de coleta de dados, meta-informação no contexto e propósito. Assim, a padronização de dados é a chave para garantir a qualidade dos dados. A falta de padronização pode criar os chamados dados incorretos que podem resultar na perda da confiabilidade dos modelos de análise. (WOOL et al., 2018).

Além desses impactos diretos, dados de baixa qualidade também podem levar a muitos efeitos intangíveis como insatisfação do cliente, tomada de decisão ineficaz e uma capacidade reduzida de executar estratégia (LAMBA; SINGH, 2018)

5. Conclusões

As perspectivas para o futuro da pesquisa em qualidade de *Big Data* são convergentes entre os autores citados. Acredita-se que novos estudos vão incorporar os conhecimentos já existentes e estabelecer sistemas de aprendizagem orientados por dados que envolvam tanto a qualidade quanto as interações humanas. O processamento destes dados em grande escala possui uma grande promessa em estabelecer soluções técnicas mais eficientes e regimes operacionais

novos. Novos modelos estatísticos e computacionais são necessários como ferramentas para melhorar tal abordagem.

Por outro lado, as várias fontes de erro que perturbam os processos analíticos devem ser melhor compreendidas para que se facilitem as escolhas de gerentes e tomadores de decisões, quando na implantação e utilização destas tecnologias. Neste sentido, é necessário estudar como integrar dados de diferentes fontes e garantir a velocidade de transmissão em tempo real, que representa a base das demais tecnologias 4.0. O papel da qualidade de dados se torna cada vez mais importante na utilização da tecnologia de *Big Data*, e as generalizações devem ser cautelosas.

Assim, existe uma escolha crítica a ser feita pelos gestores entre esforço de investimento em melhoria de qualidade de dados e o aumento no volume de dados. Já é notável que as empresas que gerenciam seus dados como recurso estratégico e investem em qualidade avançam em confiabilidade, gerando valor tangível ao cliente. Portanto, para manter a competitividade, as empresas devem investir em serviços específicos para *Big Data*, o que requer treinamento e desenvolvimento da equipe e do sistema de automação.

Referências

ABDULLAH, N.; ISMAIL, S. A.; SOPHIAYATI, S.; SAM, S. M. *Data Quality in Big Data: A Review*. Int. J. Adv. Soft. Comp.. Appl. Vol. 7, n. 3, 2015.

ALJUMAILI, M.; WANDT, K.; KARIM, R.; TRETEN, P. "eMaintenance ontologies for data quality support". *Journal of Quality in Maintenance Engineering*. Vol. 21, p. 358-374, 2015.

BENHENNI, A. L. *Pragmatic Big Data and smart manufacturing*. IN: 18th International Congress of Metrology, 2017.

BERG, H.; WILTS, H. *Digital platforms as market places for the circular economy—requirements and challenges*. IN: Management Forum, 2018.

BIANCONE, P. P.; SECINARO, S.; BRESCIA, V. *A Review of Big Data Quality and an Assessment Method and features of Data Quality for Public Health Information Systems*. International Journal of Management Sciences and Business Research. Vol. 7, n. 1, p. 2226-8235, 2018.

BOKRANTZ, J.; SKOOGH, A.; LÄMKULL, D.; HANNA, A.; PERERA, T. *Data quality problems in discrete event simulation of manufacturing operations*. Simulation: Transactions of the Society for Modeling and Simulation International. Vol. 0, n. 0, 2017.

CASALE, G. et al. *DICE: Quality-Driven Development of Data-Intensive Cloud Applications*. IN: IEEE/ACM 7th International Workshop on Modeling in Software Engineering, 2015.

CLARKE, R. *Big Data, big risks*. Information Systems Journal. Vol. 26, p. 77–90, 2016.

DING, Z.; MEI, G.; CUOMO, S.; LI, Y.; XU, N. *Comparison of Estimating Missing Values in IoT Time Series Data Using Different Interpolation Algorithms*. International Journal of Parallel Programming. 2018.

FAGUNDES, P. B.; MACEDO, D. D. J.; FREUND, G. P. *A produção científica sobre qualidade de dados em Big Data: um estudo na base de dados web of Science*. Revista Digital de Biblioteconomia e Ciência da Informação. Vol.16, n.1, p. 194-210, 2018.

FIRMANI, D.; MACELLA, M.; SCANNAPIECO, M.; BATINI, C. *On the Meaningfulness of “Big Data Quality” (Invited Paper)*. Data Science Engineering. Vol. 1, n. 1, p. 6-20, 2016.

GALLAGHER, C. V.; LEAHY, K.; O'DONOVAN, P. BRUTON, K. O'SULLIVAN, D. T. J. *Development and application of a machine learning supported methodology for measurement and verification (M&V) 2.0*. Energy & Buildings. Vol. 167, p. 8-22, 2018.

GANDOMI, A.; HAIDER, M. *Beyond the hype: Big Data concepts, methods, and analytics*. International Journal of Information Management. Vol. 35, n. 2, p. 137–144, 2015.

GE, M.; BANGUI, H.; BUHNOVA, B. *Big Data for Internet of Things: A Survey*. Future Generation Computer Systems. Vol. 87, p. 601–614, 2018.

GRUSCHKA, N.; LÜSSEM, J. *Information Quality Challenges in Industry 4.0*. IN: ICIQ 2016, Ciudad Real (Spain), Paper 10, 2016.

HAUG, A.; ARLBJØRN, J. S. *Barriers to master data quality*. Journal of Enterprise Information Management Vol. 2, n. 3, p. 288-303, 2011.

HAZEN, B.T.; BOONE, C. A.; EZELL, J.D.; et al. *Data Quality for data science, predictive analysis, and Big Data in supply chain management: an introduction to the problem and suggestions for research and applications*. International Journal of Production Economy. Vol. 154, p. 72–80, 2014.

HEINRICH, B.; HRISTOVA, D.; Klier, M.; SCHILLER, A.; SZUBARTOWICZ, M. *Requirements for Data Quality Metrics*. Journal of Data and Information Quality. Vol. 9, n. 2, 2018.

HÖLSCHER, D.; BAYER, T.; RUF, P.; REICH, C.; GUTY, F. *A Big Data Quality Preprocessing and Domain Analysis Provisioner Framework using Cloud Infrastructures* IN: ALLDATA 2018: The Fourth International Conference on Big Data, Small Data, Linked Data and Open Data. 2018.

KAMDI, A. *Big Data Quality Assurance and Testing Framework*. International Journal for Innovative Research in Science & Technology. Vol. 4, n. 10, 2018.

KHODABAKHSH, A.; ARI, I.; BAKIR, M. *Cloud-based Fault Detection and Classification for Oil & Gas Industry*. IN: A. JORGE, G. LARRAZABAL, P. GUILLEN, R.L. LOPES (eds.): Proceedings of the Workshop on Data Mining for Oil and Gas (DM4OG), Houston, Texas, USA, 2017.

LAI, ST. *An Iterative and Incremental Data Quality Improvement Procedure for Reducing the Risk of Big Data Project*. Journal of Software. Vol. 12, n. 12, 2017.

LAMBA, K.; SINGH, S. P. *Modeling Big Data enablers for operations and supply chain management*. The International Journal of Logistics Management. Vol. 29, n. 2, pp. 629-658, 2018.

LANEY, D. *3D Data Management: Controlling Data Volume, Velocity, and Variety*. Application Delivery Strategies. Arquivo 949, 2001.

LEE, J.; KAO, AH.; YANG, S. *"Service innovation and smart analytics for industry 4.0 and Big Data environment"*. Procedia CIRP. Vol. 16, p. 3-8, 2014.

LI, X.; GONG, X.; LIU, Y. *Development and application of basis database for materials life cycle assessment in china*. IN: 17th IUMRS International Conference in Asia (IUMRS-ICA), 2016.

MADHIKERM, M. KUBLER, S.; ROBERT, J.; BUDA, A.; FRÄMLING, K. *Data quality assessment of maintenance reporting procedures*. Expert Systems With Applications. Vol. 63, p. 145–164, 2016.

MAZAHUA, L. R. et al. *A general perspective of Big Data: applications, tools, challenges and trends*. J Supercomput. Vol 72. P. 3073–3113, 2016.

MITTAL, A. *Trustworthiness of Big Data*. International Journal of Computer Applications. Vol. 8, n.9, p. 0975 – 8887, 2013.

MOYNE, J.; ISKANDAR, J. *Big Data Analytics for Smart Manufacturing: Case Studies in Semiconductor Manufacturing*. Processes, Vol. 5, n. 9, 2017.

NGUYEN, T.; ZHOU, L.; SPIEGLER, B.; IEROMONACHOU, P. LIN, Y. *Big Data analytics in supply chain management: A state-of-the-art literature review*. Computers and Operations Research. Vol. 98, p. 254–264, 2018.

PAWAR, S. H.; THAKORE, D. *An Assessment Model to Evaluate Quality Attributes In Big Data Quality*. International Journal of Computer Science Trends and Technology. Vol. 5, n. 2, 2017.

REIS, M. S.; KENETT, R. *Assessing the Value of Information of Data-Centric Activities in the Chemical Processing Industry 4.0*. Process Systems Engineering, Vol. 00, n. 0, 2018.

RØDSETH, O. J.; PERERA, L. P.; MO, B. *Big Data in Shipping - Challenges and Opportunities*. IN: 15th International Conference on Computer and IT Applications in the Maritime Industries – COMPIT, 2016.

SADIQ, S. et al. *Data Quality – The Role of Empiricism*. SIGMOD Record. Vol. 46, n. 4, 2017.

SANTOS, M.Y.; SÁ, J.O.; COSTA, C.; GALVÃO, J.; ANDRADE, C.; MARTINHO, B.; LIMA, F.V.; COSTA, E. et al. *A Big Data analytics architecture for Industry 4.0*. IN: World Conference on Information Systems and Technologies. p.175–184, 2017.

SANTOS, M. Y.; OLIVEIRA E SÁ, J.; ANDRADE, C.; LIMA, F. V.; COSTA, E.; COSTA, C.; MARTINHO, B.; GALVÃO, J. *A Big Data system supporting Bosch Braga Industry 4.0 strategy*. International Journal of Information Management. Vol. 37, p. 750–760, 2017.

SONG, Z.; SUN, Y.; WAN, J.; LIANG, P. *Data quality management for service-oriented manufacturing cyber-physical systems*. Computers and Electrical Engineering. Vol. 64, p. 34-44, 2017.

THOBEN, KD.; WIESNER, S. *“Industrie 4.0” and Smart Manufacturing – A Review of Research Issues and Application Examples*. International Journal. of Automation Technology. Vol.11, n.1, 2017.

UTZ, F.; NEUMANN, C.; TAFRESCHI, O. *How to Discover Knowledge for Improving Availability in the Manufacturing Domain?* IN: Proceedings of the 51st Hawaii International Conference on System Sciences, 2018.

VERMA, S.; BHATTACHARYYA, S. S.; KUMAR, S. *An extension of the technology acceptance model in the Big Data analytics system implementation environment*. Information Processing and Management. Vol. 54, p. 791–806, 2018.

WOODALL, P.; BOREK, A.; GAO, J.; OBERHOFER, M.; KORONIOS, A. *An Investigation of How Data Quality is Affected by Dataset Size in the Context of Big Data Analytics*. IN: 19th International Conference on Information Quality (ICIQ), Xi'an, 2014.

WOOL, J.; SHIN, SJ.; SEO, W.; MEILANITASARI, P. *Developing a Big Data Analytics Platform for Increasing Sustainability Performance in Machining Operations*. The International Journal of Advanced Manufacturing Technology. 2018.

YAO, L.; GE, Z. *Big Data quality prediction in the process industry: A distributedparallel modeling framework*. Journal of Process Control. Vol. 68, p. 1-13. 2018.

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