# Mecanismos tecnológicos de segurança da informação no tratamento da veracidade dos dados em ambientes *Big Data*

***Gislaine Parra Freund*** [I]

***Priscila Basto Fagundes*** [II]

***Douglas Dyllon Jerônimo de Macedo*** [III]

***Moisés Lima Dutra*** [IV]

[I] *Universidade Federal de Santa Catarina, SC, Brasil. Mestre em Ciência da Computação*.

[II] *Doutoranda em Ciência da Informação.*

[III] *Professor do Programa de Pós-graduação em Ciência da Informação.*

[IV] *Professor do Programa de Pós-graduação em Ciência da Informação.*

***http://dx.doi.org/10.1590/1981-5344/3348***

*Com o fenômeno Big Data, o volume de dados e informações que são geradas, processadas, armazenadas e utilizadas para a tomada de decisões aumenta em uma velocidade expressiva, e um dos grandes desafios é garantir que estes dados e informações sejam e permaneçam confiáveis. Uma das características relacionadas aos ambientes Big Data é denominada veracidade e diz respeito a qualidade e credibilidade dos dados, uma vez que é fundamental que o dado seja confiável para produzir informações confiáveis. Este artigo tem como objetivo realizar uma análise dos mecanismos tecnológicos de segurança da informação, identificando aqueles que contribuem para o aumento da veracidade dos dados em ambientes de Big Data. Como resultado deste estudo, foi identificado um conjunto de requisitos relacionados com a veracidade dos dados e para cada um desses requisitos foi proposto o uso de mecanismos de segurança da informação. Este artigo pretende contribuir com novas pesquisas sobre o tema segurança da informação dentro da Ciência da Informação, uma vez que apresentará uma reflexão sobre os mecanismos*

*disponíveis para tratar a segurança da informação, contribuindo com o aumento da confiabilidade das informações geradas a partir de ambientes que envolvam grandes quantidade de dados.*

***Palavras Chave:*** *Big Data, Veracidade, Segurança da Informação.*

# Technological information security mechanisms in the treatment of data veracity in Big Data environments

*With the Big Data phenomenon, the volume of data and information that is generated, processed, stored and used for decision making is increasing at an expressive speed and one of the great challenges is to ensure that this data and information is and remains reliable. One of the characteristics related to Big Data environments is called veracity and refers to the quality and credibility of the data, since it is fundamental that the data is reliable to produce reliable information. This article aims to perform an analysis of the technological mechanisms of information security, identifying those that contribute to increase the accuracy of data in Big Data environments. As a result of this study, a set of requirements related to the veracity of the data was identified and for each of these requirements the use of information security mechanisms was proposed. This article intends to contribute with new research on the subject of Information Security within Information Science, since it will present a reflection on the mechanisms available to treat information security, contributing to the increase of the reliability of information generated from environments that involve large amounts of data.*

***Keywords:*** *Big Data, Veracity, Information Security.*

Recebido em 17.11.2017 Aceito em 02.04.2019

## 1 Introdução

A utilização de dados e informações para a tomada de decisão e identificação de novas oportunidades é algo praticado pelas organizações e objeto de estudo em diversas áreas do conhecimento, tais como:

Ciência da Informação, Saúde, Educação, Engenharias, Ciência da Computação, entre outras. E cada vez mais a quantidade de dados e a velocidade em que são gerados, contribuem com o aumento da complexidade dos ambientes tecnológicos necessários para lidar com os mesmos. Se por um lado, a qualidade dos dados e informações disponíveis para análise é importante para a obtenção de resultados mais precisos, por outro, o grande desafio é garantir que estes dados e informações sejam confiáveis.

A conscientização dos problemas relacionados à falta de confiabilidade dos dados e consequentemente das informações geradas a partir deles, vem crescendo substancialmente nos últimos anos. Uma pesquisa envolvendo mais de 140 empresas de vários setores e países evidenciou que decisões baseadas em dados inúteis resultam em perdas financeiras significativas. Nessa pesquisa foi solicitado aos entrevistados que estimassem o impacto financeiro gerado pela falta de confiabilidade das informações em suas organizações e os resultados foram perdas anuais médias estimadas em US$ 8,2 milhões, sendo que, algumas organizações indicaram perdas de até US$ 100 milhões em um ano (BAŠKARADA; KORONIOSA, 2014). E dentro desse contexto é possível observar que o desafio, de prover, usar e disponibilizar dados e informações de forma confiável, está inserido nos diferentes segmentos da sociedade, seja como provedor de informações ou mesmo como usuário delas.

Saracevic (1996) afirma que dentre os objetivos da Ciência da Informação está o de fornecer meios para a disponibilização de informações relevantes para indivíduos, grupos e organizações envolvidas com a ciência e tecnologia. E coadunando com este pensamento Moraes (2010) conceitua segurança da informação como um conjunto de medidas que visa proteger e preservar informações e sistemas de informação. O referido autor afirma ainda, que toda e qualquer informação deve ser correta, precisa e estar disponível para ser armazenada, recuperada, processada e disponibilizada de forma segura e confiável. Em concordância com essa linha de raciocínio está Neto (2016), que conceitua *Big Data* como ambientes que possuem grandes volumes de dados, e que se os dados processados não forem autênticos, a informação gerada não será confiável. Além do volume, possível encontrar na literatura outras características relacionadas a este tipo de ambiente, entre elas está a velocidade, variedade, veracidade, valor, variabilidade e visualização (ZIKOPOULOS *et al*., 2012; GANDOMI; HAIDER, 2015; DEVAN, 2016).

A fim de contribuir para o aumento da confiabilidade dos dados e informações geradas nos ambientes *Big Data*, este estudo se propõe a realizar uma análise com foco na característica veracidade em *Big Data***,** apresentando as relações deste aspecto com os mecanismos tecnológicos de segurança da informação.

## 2 A segurança da informação e seus mecanismos

A segurança da informação transcende os controles computacionais permitindo assim que diferentes abordagens sejam utilizadas em sua definição conceitual. Moraes (2010) define segurança da informação como um processo para proteger informações do mau uso, intencional ou não, realizados por pessoas internas ou externas à organização. Para Ferreira (2003), a segurança da informação possibilita a utilização dos recursos que suportam as informações necessárias para as atividades estratégicas, táticas e operacionais de maneira confiável de uma organização. E Sêmola (2003) conceitua segurança da informação como uma área do conhecimento dedicada a proteção dos ativos de informação contra acessos não autorizados, alterações indevidas ou sua indisponibilidade. A segurança da informação é alcançada com a implementação de um conjunto adequado de controles, incluindo políticas, processos, procedimentos, estrutura organizacional e funções de software e hardware (ISO/IEC 2002:2013).

Em suas origens e por questões históricas, a segurança da informação geralmente é associada ao princípio da confidencialidade/sigilo onde controles de segurança em uma versão mais rudimentar eram adotados para garantir o sigilo de informações críticas (SINGH, 2005). A confidencialidade é um princípio importante dentro da segurança da informação, porém existem outros que devem ser considerados para garantir proteção efetiva dos dados e informações. Sêmola (2003) afirma que são três os princípios básicos que norteiam a implementação da segurança da informação, são eles, confidencialidade, integridade e disponibilidade. Para Beal (2005), informações críticas devem ser protegidas para evitar sua destruição, indisponibilidade temporária, adulteração ou divulgação não autorizada. Fontes (2012) também descreve que o grau de disponibilidade, integridade e confidencialidade protegerá a informação para que a organização operacionalize os seus negócios e atenda a seus objetivos.

Nakamur e Geus (2007) apresentam mais dois princípios adicionais para que a proteção da informação seja alcançada, a autenticidade e o não-repúdio - também encontrada na literatura como irretratabilidade. Já Demétrio (2003) apresenta também o princípio da tempestividade que faz referência aos documentos eletrônicos, para que tenham as mesmas garantias de confiabilidade existentes nos documentos em papel tais como: autenticidade, integridade e tempestividade. Para o desenvolvimento deste estudo são considerados os seis princípios da segurança da informação apresentados por estes autores, observando as similaridades de nomenclatura e de interpretação existente entre eles:

- Confidencialidade: é o princípio que garante que a informação seja acessada somente por entidades[1] autorizadas.
- Integridade: assegura que a informação é íntegra e fiel, ou seja, que ela não sofra alterações por entidades não autorizadas pelo seu proprietário.
- Disponibilidade: trata do princípio que assegura que a informação esteja disponível sempre que for necessária.
- Autenticidade: garante que as entidades envolvidas em um processo contendo informações digitais são autênticas, ou seja, se são verdadeiras garantindo que a informação é proveniente da fonte anunciada.
- Irretratabilidade: também denominada como não-repúdio, refere-se a garantia que uma entidade não negue a autoria de algo realizado por ela.
- Tempestividade: é o princípio que garante a validade de uma informação digital ao longo do tempo.

Para cada um dos princípios apresentados, são adotados mecanismos tecnológicos específicos que são as ferramentas, tecnologias e protocolos utilizados no processo de segurança da informação (NAKAMURA; GEUS, 2007).

Quadro 1. Mecanismos tecnológicos da segurança da informação

| MECANISMO DE SEGURANÇA | DESCRIÇÃO |
|---|---|
| **Criptografia** | Possui importância fundamental para a segurança da informação, uma vez que é a base para diversas tecnologias e protocolos utilizados com objetivo de garantir confidencialidade, integridade, autenticação e irretratabilidade das informações. Este mecanismo transforma dados legíveis em ilegíveis utilizando um código de maneira que somente entidades autorizadas e detentoras do mesmo conseguem descriptografá-los e interpretá-los (NAKAMURA; GEUS, 2007). |
| ***Hashing*** | São cálculos matemáticos utilizados em algoritmos que produzem o histórico da informação possibilitando identificar se a mesma foi alterada. Algoritmos de cálculo de *hashing* são usados para garantir a integridade e identificar se ocorreram mudanças não previstas. (MORAES, 2010). |
| **Assinatura digital** | É a combinação de mecanismos de *hashing* e criptografia, utilizada para garantir a autenticidade, a integridade e a irretrabilidade da informação (MORAES, 2010). |
| **Controle de acesso** | Trata da limitação de acesso às informações e deve ser implementado considerando a "necessidade de conhecer" e a "necessidade de acesso". A norma recomenda que as permissões de |

[1] Entende-se por entidades, pessoas ou equipamentos tecnológicos participantes de uma transação.

| | |
|---|---|
| | acesso sejam aprovadas pelo responsável pela informação. Além disso, o recurso de "perfil" pode ser adotado para autorizar não somente os acessos, mas também as ações individuais ou de um grupo de usuários (ISO/IEC 2002:2013). |
| ***Backup*** | São cópias de segurança que garantem a recuperação das informações em caso de perda ou indisponibilidade das mesmas em suas bases originais (ISO/IEC 2002:2013). |
| **Certificados Digitais** | Materializam o uso da assinatura digital e possibilitam o uso da criptografia, sendo emitidos por autoridades certificadoras que atestam que as informações utilizadas em sua geração são verdadeiras e válidas por um determinado tempo. Com o uso de funções matemáticas é possível se obter garantia da autenticidade, irretratabilidade, integridade e confidencialidade (FONTES, 2008). |
| **Carimbo de tempo** | Garante a validade de uma informação assinada digitalmente ao longo do tempo. É um selo que atesta a data e a hora que um documento foi assinado digitalmente assegurando que o mesmo não foi adulterado no intervalo de tempo entre a assinatura e a consulta ao documento. Este mecanismo agrega uma âncora temporal ao documento eletrônico de forma que algumas características presentes em documentos físicos, como identificação de autoria e alteração no documento de forma imperceptível, também estejam presentes em documentos eletrônicos para evitar possíveis contestações jurídicas (DEMÉTRIO, 2003). |

Fonte: Elaborado pelos autores

Os mecanismos apresentados no Quadro 1, são os mecanismos tecnológicos de segurança da informação que serão utilizados na análise que se propõe este artigo. Vale ressaltar que certificados digitais e carimbo do tempo são tecnologias que implementam os mecanismos de segurança, porém são abordadas neste trabalho por se tratarem de tecnologias consolidadas no âmbito jurídico.

# 3 Big Data

Para Mcafee e Brynjolfsson (2012), o termo *Big Data* faz referência a grandes volumes de dados que possuem diferentes características, são heterogêneos e que se originam de diferentes fontes. Em uma realidade no qual as organizações estão gerando enormes quantidades de dados, o que demanda um processo de gestão específico para garantir a sua qualidade, as soluções e práticas de *Big Data* se fazem necessárias quando as tecnologias e técnicas tradicionais não são suficientes para a execução de atividades relacionadas ao gerenciamento de grandes *datasets*[2] (ERL; KHATTAK; BUHLER, 2016).

Outra questão a ser considerada, é a quantidade de atores envolvidos na geração de dados e informações, uma vez que, com o

[2] *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).

advento da internet, das redes sociais e dispositivos móveis, este número aumentou consideravelmente. Atualmente este cenário se ampliou ainda mais com a progressiva implantação da Internet das Coisas, proposta que pressupõe a interconexão de todas as "coisas", gerando informações de eventos e transações realizadas e capturadas pelas mesmas. Mayer e Cukier (2013) reforçam esta observação ao citar que a adoção massiva de telefones celulares pela sociedade e o grande número de computadores e sistemas de informação existentes, gera uma avalanche de informações e complementa que a rapidez com o qual esses dados são gerados e acumulados é um fator importante para definir *Big Data*.

Davenport (2014) corrobora com esse ponto de vista, ao relatar que Big Data está inegavelmente relacionado ao grande volume de dados, porém acrescenta que esta definição não é completa. O autor afirma que enquanto o fator volume recebe toda a atenção, um dos aspectos mais desafiador do Big Data está relacionado com a falta de estrutura para lidar com todos estes dados. Ambientes com servidores únicos, bancos de dados estruturados em linhas e colunas e repositórios estáticos precisam estar adequados para armazenar grandes quantidades de dados, estando estes estruturados ou não, sendo de diferentes tipos e formatos, gerando assim um fluxo intenso e contínuo.

De acordo com Fagundes, Macedo e Parra (2018), existem na literatura, diferentes pontos de vista em relação as características que compõem *Big Data.* O que parece consenso é que, para que um conjunto de dados seja considerado como tal, ele deve possuir no mínimo uma dessas particularidades. Três delas foram inicialmente identificadas por Doug Laney no início de 2001, são elas:

- Volume: diz respeito a grande quantidade de dados gerada por organizações, usuários e dispositivos;
- Velocidade: está relacionada com o tempo de resposta para determinada requisição.
- Variedade: refere-se aos diversos tipos e formatos de dados que são gerados e precisam ser suportados pelos ambientes de *Big Data*;

Alguns anos depois outras características foram adicionadas ao conjunto de aspectos relacionados a B*ig Data* e são elas: veracidade, valor, variabilidade e visualização (ZIKOPOULOS *et al*., 2012; GANDOMI; HAIDER, 2015), onde:

- Veracidade: está relacionada com a qualidade e fidelidade dos dados, ou seja, com o grau de precisão e confiabilidade que o dado possui;
- Valor: refere-se à utilidade dos dados e a sua importância dentro de um determinado contexto.
- Variabilidade: é a mudança de significado que o dado sofre ao longo do tempo;

- Visualização: refere-se à eficácia da forma de apresentação dos dados.

## *4* Veracidade em *Big Data*

A característica veracidade em ambientes *Big Data*, refere-se ao grau de credibilidade dos dados, sendo que os mesmos devem apresentar confiabilidade significativa para proporcionar valor e utilidade aos resultados gerados a partir deles (LUKOIANOVA; RUBIN, 2014). Na concepção das autoras os dados precisam ser avaliados quanto a sua veracidade, objetividade e credibilidade para garantir a produção de informações verdadeiras, objetivas e credíveis.

Walker (2012) afirma que é necessário repensar as arquiteturas tradicionais de repositórios para que estas sejam preparadas para receber e processar grandes volumes de dados estruturados e não-estruturados. Segundo o autor, os dados não-estruturados trazem por natureza uma quantidade significativa de imprecisão e incerteza, a exemplo dos dados originados a partir de mídias sociais, os quais são inerentemente imprecisos. Os níveis de imprecisão e incerteza dos dados podem variar de uma base de dados para outra e as decisões precisam ser tomadas sob os dados que apresentam maior nível de veracidade.

Em relação aos requisitos para verificar a veracidade dos dados, Claverie-Berge (2012) propõe a inconsistência que se refere aos dados divergentes sobre um mesmo fato; a incompletude que está relacionada a falta de dados essenciais para o atingimento de um determinado objetivo; a ambiguidade que diz respeito a possibilidade de ocorrer interpretações distintas e equivocadas dos dados; a latência que está relacionada com o tempo em que o dado é coletado até o momento em que este gera algum tipo de resultado; e por fim os modelos de aproximação que são algoritmos que consideram a correlação entre os dados para tratá-los.

Sordi, Meireles e Grijo (2008) trazem a confiabilidade dentro do contexto de dimensão para avaliar a qualidade dos dados, a qual se entende ser um viés complementar aos requisitos de veracidade propostos por Claverie-Berge (2012). A confiabilidade dos dados está diretamente ligada à percepção do usuário quanto a autoridade e confiabilidade da fonte e do conteúdo. A confiabilidade não assegura que o dado é verdadeiro, mas contribui para estabelecer o quanto ele é confiável por ter sido gerado por uma fonte confiável.

Sob o ponto de vista de Lukoianova e Rubin (2014, o tema veracidade em Big Data está relacionado com o gerenciamento de incertezas. As autoras apresentam uma proposta para a redução das incertezas quanto ao conteúdo de dados textuais, associando ferramentas de linguística computacional que podem ser usadas para medir três dimensões de veracidade: a objetividade que está relacionada com a forma particular de escrita de quem gera a informação; a veracidade que se refere ao grau de verdade existente na informação; e a credibilidade que diz respeito a quanto da informação é crível.

Conforme será apresentado na próxima sessão, é possível observar uma escassez de estudos a respeito dos requisitos de veracidade dos dados, visto que e os autores citados possuem diferentes abordagens para tratar sobre o assunto. Alguns com entendimentos similares e outros complementares, porém, todos com o mesmo objetivo – obtenção de maior precisão e credibilidade dos dados.

# 5 Procedimentos metodológicos

Quanto aos procedimentos metodológicos utilizados na escrita deste artigo, pode-se dizer que o mesmo possui uma abordagem qualitativa, visto que não se preocupa com representatividade numérica, e sim com o aprofundamento da compreensão de um grupo social, de uma organização, etc. Quanto à natureza, é uma pesquisa básica pois objetiva gerar conhecimentos novos, úteis para o avanço da ciência, sem aplicação prática prevista (GERHARDT; SILVEIRA, 2009).

Quanto aos objetivos, é uma pesquisa de caráter exploratório devido ao fato de ter como objetivo um melhor entendimento do problema a ser estudado, possuindo o propósito de promover uma maior familiaridade com os temas para torná-los mais explícitos e também auxiliar na construção de novas hipóteses (GIL, 2002). Quanto aos procedimentos pode ser classificada como uma pesquisa bibliográfica, uma vez que foi elaborada a partir do levantamento de referências teóricas já analisadas e publicadas em meios escritos e eletrônicos, como livros, artigos científicos e sites na internet (FONSECA, 2002).

A fim de verificar a originalidade e possível contribuição deste estudo para a área da Ciência da Informação foram realizadas consultas às bases de dados: LISA (*Library & Information Science Abstracts*), LISTA (*Library, Information Science & Technology Abstracts*), SciELO (*Scientific Eletronic Library Online*) e *Web of Science*. As buscas nestas bases foram realizadas entre os dias 15/05/2017 e 18/05/2017 e restringiram-se as publicações na área da Ciência da Informação, sem limitações para o período de publicação das mesmas. Por se tratarem de bases de dados internacionais, optou-se pela utilização dos termos na língua inglesa. Para os termos "*Big Data*" *and Veracity* foram encontradas 27 publicações, para os termos "*Big Data*" *and* "*Information Security*", 7 publicações e por fim para os termos que contemplavam os três assuntos tratados neste estudo, "*Big Data*" *and* "*Veracity*" *and* "*Information Security*", o resultado foi nenhuma publicação. Sendo assim, diante dos números apresentados, nota-se que ainda são incipientes os estudos na área de *Big Data* sob a perspectiva da veracidade relacionada com a segurança da informação.

# 6 Análise dos resultados

A fim de selecionar os requisitos relacionados com a veracidade dos dados para fazerem parte deste estudo, foi realizada uma análise dos requisitos apresentados na sessão que trata da veracidade em Big Data de acordo com os autores Claverie-Berge (2012); Sordi, Meireles e Grijo

(2008) e Lukoianova e Rubin (2014). A análise considerou a complementaridade, as similaridades e equivalências existentes entre eles. A partir dessa análise, os requisitos foram compilados e selecionados para a formação do conjunto de requisitos a ser trabalhado neste artigo. O Quadro 2 apresenta a seleção dos requisitos obtidos na análise a partir da literatura.

Quadro 2: Seleção dos requisitos de veracidade apresentados na literatura.

| AUTOR | REQUISITOS SELECIONADOS | REQUISITOS EXCLUÍDOS |
|---|---|---|
| Claverie-Berge (2012) | inconsistência, incompletude, ambiguidade, latência e modelos de aproximação | - |
| Sordi, Meireles e Grijo (2008) | confiabilidade | - |
| Lukoianova e Rubin (2014) | veracidade | objetividade e credibilidade |

Fonte: Elaborado pelos autores

Conforme pode ser observado, os requisitos propostos por Claverie-Berge (2012) e Sordi, Meireles e Grijo (2008) serão utilizados na íntegra. A seleção do requisito veracidade proposto por Lukoianova e Rubin (2014) se justifica por se entender que no contexto apresentado pelas autoras, o mesmo é complementar ao requisito de confiabilidade sugerido por Sordi, Meireles e Grijo (2008). Os requisitos objetividade e credibilidade foram excluídos, por se acreditar que a objetividade trata de um fator especificamente de linguagem textual e a credibilidade possui o conceito similar a confiabilidade apresentada por Sordi, Meireles e Grijo (2008), já contemplada no conjunto de requisitos.

Assim, apresentam-se sete requisitos relacionados com a veracidade dos dados obtidos a partir da literatura apresentada. A partir das análises empreendidas neste trabalho, além destes, decidiu-se pela inclusão de outros dois requisitos entendidos como importantes para atender os objetivos deste estudo, são eles:

- Legalidade: no Brasil existem algumas legislações específicas para a proteção e uso de dados, tais como a Lei 12.965 de 2014 sobre o Marco Civil da Internet, o Projeto de Lei 5276/2016 sobre proteção de dados pessoais e a Lei 9.279 de 1996 sobre proteção da propriedade intelectual, entre outras. Estas, assim como outras legislações aplicadas em outros países, são importantes no contexto de *Big Data*, tanto para o acesso, quanto para o uso e disponibilização dos dados. Este requisito foi incluído por se entender que a legislação pertinente a cada situação, aplicável às localidades às quais os dados estão sendo acessados, utilizados e disponibilizados, precisa ser

atendida para transmitir aos usuários do *Big Data* a confiança também sob a perspectiva legal.

- Privacidade: refere-se à proteção de dados privados, referindo-se aos seu uso, armazenamento e disponibilização. É importante considerar quais dados podem ser obtidos e mantidos para que seja oferecida a privacidade adequada a cada tipo e assim proporcionar aos usuários destes ambientes a confiança de que os dados privados serão tratados como tal, desde a sua obtenção até a sua disponibilização. É importante considerar ainda, que dados publicados em diferentes mídias podem ser visualizados publicamente e serem utilizados para gerar conclusões errôneas, por se tratarem de dados isolados e fora de contexto. Porém, o requisito de privacidade sob este ponto de vista não é escopo de estudo deste trabalho.

Sendo assim, os requisitos referentes à veracidade dos dados em ambientes de *Big Data* a serem analisado neste trabalho totalizou em nove itens, assim definidos: inconsistência, incompletude, ambiguidade, latência, modelos de aproximação, confiabilidade, veracidade, legalidade e privacidade.

A seguir serão apresentados os resultados das análises sobre cada um dos requisitos, os mecanismos de segurança que podem contribuir para o seu tratamento e as justificativas para as suas proposições.

Quadro 3: Mecanismos de segurança da informação para o requisito inconsistência

| REQUISITO | MECANISMOS |
|---|---|
| **INCONSISTÊNCIA** | - Certificado digital nas fontes de dados<br>- Mecanismos de validação de dados de entrada nos campos relevantes, no sistema de armazenamento |

Fonte: Elaborado pelos autores

A proposição dos mecanismos de segurança para o requisito inconsistência apresentado no Quadro 3, considera dois aspectos, a inconsistência gerada na fonte dos dados (no caso de fontes não confiáveis), onde o mecanismo proposto é o uso de certificado digital emitido por uma Infraestrutura de Chaves Públicas (ICP) de acreditação internacional nas fontes de dados, para que seja comprovada sua autenticidade. Este mecanismo atesta que a base não é falsa, pelo fato de ter sido certificada por um órgão regulador de certificados digitais autorizado e reconhecido. E a inconsistência na base de armazenamento dos dados coletados, onde a proposição é a adoção de mecanismos de validação dos dados de entrada nos campos relevantes do sistema. Esta proposta considera a possibilidade de haver uma falha de segurança que

permita que dados inconsistentes (que não condizem com a realidade) sejam registrados nos sistemas.

Quadro 4: Mecanismos de segurança da informação para o requisito incompletude

| REQUISITO | MECANISMOS |
|---|---|
| **INCOMPLETUDE** | - Uso de hashing no trafego dos dados.<br>- Uso de hashing no armazenamento dos dados.<br>- Autenticação para acesso à base dos dados coletados.<br>- Adoção do recurso de perfil de acesso aos dados.<br>- Adoção de sistemas de replicação de dados. |

Fonte: Elaborado pelos autores

Conforme apresentado no Quadro 4, os mecanismos de segurança para o requisito incompletude se referem a adoção do mecanismo de *hashing* no tráfego que permite que seja verificado se os dados recebidos são os mesmos extraídos da fonte, ou seja, afere se os dados recebidos são fiéis aos originais (fonte). A completude dos dados também deve ser mantida durante o período em que os mesmos permaneçam armazenados. Desta forma, é possível identificar a existência de alterações nos dados e os acessos são segmentados conforme as necessidades – com as opções "somente leitura" dos dados ou "acesso total". Esses mecanismos reduzem os riscos de exclusões e alterações dos dados por entidades não autorizadas e, caso ocorram, que sejam identificadas.

A replicação dos dados coletados também pode ser adotada para promover o "*backup*" dos dados e garantir sua completeza, no caso de indisponibilidade de servidores de armazenamento. Além disso, os sistemas de replicação também utilizam os mecanismos de *hashing* para assegurar que os dados permaneçam completos após sua replicação e integração. Vale ressaltar que não foram identificados mecanismos de segurança para tratar a incompletude quando o dado for gerado incompleto e se houver falhas no mecanismo de coleta dos mesmos. Os mecanismos apresentados aqui visam assegurar a completude durante o tráfego e armazenamento dos dados.

Quadro 5: Mecanismos de segurança da informação para o requisito ambiguidade

| REQUISITO | MECANISMOS |
|---|---|
| **AMBIGUIDADE** | - Para tratar a ambiguidade dos dados não foram identificados mecanismos tecnológicos de segurança da informação diretamente associados. |

Fonte: Elaborado pelos autores

A ambiguidade refere-se a imprecisão dos dados de forma que gere interpretações diferentes e equivocadas. Mesmo não se tratando de mecanismos de segurança, sugere-se que sejam adotados metadados consistentes, técnicas de semântica, como por exemplo, o uso de ontologias e dicionários de dados para minimizar as chances de resultar em falsas interpretações.

Quadro 6: Mecanismos de segurança da informação para o requisito latência

| REQUISITO | MECANISMOS |
|---|---|
| **LATÊNCIA** | - Assinatura digital.<br>- Carimbo do tempo. |

Fonte: Elaborado pelos autores

Os mecanismos de assinatura digital e carimbo do tempo podem ser utilizados para contribuir com o requisito latência, uma vez que juntos permitem aferir que mesmo os dados sendo utilizados após algum tempo após a sua coleta, os mesmos eram autênticos e válidos quando foram assinados digitalmente em sua geração (neste contexto entende-se geração, como sendo o momento em que o dado foi coletado e armazenado para posterior análise). Vale ressaltar que os mecanismos propostos para este requisito não atestam a obsolescência dos dados, mas confirmam sua validade na geração, após um período de tempo.

Quadro 7: Mecanismos de segurança da informação para o requisito modelos de aproximação

| REQUISITO | MECANISMOS |
|---|---|
| **MODELOS DE APROXIMAÇÃO** | - Para tratar os modelos de aproximação não foram identificados mecanismos de segurança da informação diretamente associados. |

Fonte: Elaborado pelos autores

Modelos de aproximação são aplicados em algoritmos para tratar os dados e identificar correlação entre eles. Para este caso, a recomendação é que os algoritmos adotem, na medida do possível, os mecanismos de segurança pertinentes, apresentados para os demais requisitos.

Quadro 8: Mecanismos de segurança da informação para o requisito confiabilidade

| REQUISITO | MECANISMOS |
|---|---|
| **CONFIABILIDADE** | - Certificado digital nas fontes de dados.<br>- Assinatura digital dos dados, na origem. |

Fonte: Elaborado pelos autores

Os dados precisam ser autênticos e assim serem percebidos por seus usuários para que sejam confiáveis. Vale destacar que, conforme Sordi, Meireles e Grijo (2008), confiabilidade não significa que a informação é verdadeira e sim o quanto ela é confiável por ter sido gerado por uma fonte confiável. Conforme apresentado no Quadro 8, propõem-se que os mecanismos de segurança para o requisito confiabilidade sejam, a adoção de fontes de dados que possuam certificado digital emitido por uma ICP com acreditação internacional, já que este mecanismo atesta que a fonte não é falsa, e a utilização de assinaturas digitais para garantir autenticidade e legitimidade dos dados na origem.

Quadro 9: Mecanismos de segurança da informação para o requisito veracidade

| REQUISITO | MECANISMOS |
|---|---|
| **VERACIDADE** | - Certificado digital nas fontes de dados.<br>- Assinatura digital dos dados, na origem. |

Fonte: Elaborado pelos autores

Para o requisito de veracidade, a proposição é o uso de certificado digital nas fontes de dados e a assinatura digital nos dados de origem. Se na confiabilidade dos dados a percepção é gerada pela confiança na fonte que gerou os dados, a veracidade também considera este fator, ou seja, considera-se que dados gerados por fontes idôneas possuem um alto grau de verdade em seu conteúdo. Sendo assim, os mecanismos propostos corroboram para isso, pois proporcionam a autenticidade e legitimidade dos dados na origem.

Quadro 10: Mecanismos de segurança da informação para o requisito legalidade

| REQUISITO | MECANISMOS |
|---|---|
| **LEGALIDADE** | - Autenticação para acesso à base dados.<br>- Adoção do recurso de perfil de acesso aos dados. |

Fonte: Elaborado pelos autores

O Quadro 10 mostra a proposição dos mecanismos de segurança para o requisito legalidade, onde foi considerada a proteção dos dados em sua disponibilização após as análises. Estes devem ser disponibilizados somente para entidades autorizadas considerando a legislação vigente e conforme o tipo de dado. O mecanismo de segurança sugerido é a autenticação com o uso de perfil para acesso às bases de dados, limitando apenas as entidades autorizadas por lei. Vale ressaltar que tão importante quanto o contexto apresentado, é observar a legalidade dos dados sob outros aspectos, tais como: legalidade na coleta de dados, para que seja apenas de fontes abertas e/ou autorizadas, o uso de dados protegidos por propriedade intelectual e outros controles como a curadoria digital para

obter um tratamento mais completo para este requisito, porém para estes casos, não foram identificados mecanismos de segurança correspondentes.

Quadro 11: Mecanismos de segurança da informação para o requisito privacidade

| REQUISITO | MECANISMOS |
|---|---|
| **PRIVACIDADE** | - Criptografia no tráfego e no armazenamento dos dados.<br>- Autenticação para acesso à base de dados. |

Fonte: Elaborado pelos autores

Para que os dados tenham sua privacidade garantida, a proposta é a adoção da criptografia, tanto no trafego quanto no seu armazenamento, desta forma é possível assegurar que somente entidades autorizadas e que possuam a chave para descriptografar os dados, tenham acesso aos mesmos. E também a adoção de autenticação para acesso à base de dados. Porém, vale ressaltar que as os mecanismos apresentados não garantem que a privacidade será respeitada em sua totalidade, pois não asseguram o cumprimento da privacidade na coleta, uso e disponibilização desses dados. O item deve ser complementado com a adoção de procedimentos e processos éticos de uso dos dados em conformidade com a lei.

Diante do apresentado, entende-se que para garantir a veracidade dos dados em ambientes de *Big Data,* os mesmos devem ser consistentes e completos, a ambiguidade, a latência e os modelos de aproximação precisam ser tratados e é fundamental que os dados sejam confiáveis, verdadeiros, e fatores relacionados a legalidade e a privacidade dos mesmos devem ser respeitados e mantidos.

# 7 Considerações finais

O termo *Big Data* é abordado na literatura sob diferentes perspectivas e não há um conceito único em relação a ele, porém a ideia principal pode ser definida como sendo ambientes que envolvam o uso de grades quantidades de dados para a tomada de decisões de forma mais precisa. De maneira geral, são sete os fatores que caracterizam *Big Data*, contudo a expectativa para a utilização desses dados é que seja baseada em dados precisos e confiáveis. Esta condição torna a característica veracidade em *Big Data* um fator indispensável para se obter valor e os resultados sejam conforme as expectativas.

Os requisitos da veracidade e mecanismos de segurança da informação foram apresentados neste estudo de forma ampla e generalista, porém, os ambientes de *Big Data* são utilizados para diferentes finalidades. Sendo assim, a aplicabilidade destes mecanismos, em menor ou maior grau, deve ser avaliada considerando as especificidade e necessidade de cada cenário.

Neste estudo, cujo objetivo foi contemplar uma análise dos mecanismos tecnológicos de segurança da informação que contribuem com a veracidade dos dados em ambientes de *Big Data,* foram apresentados os fatores que caracterizam um ambiente como *Big Data* dando ênfase ao fator veracidade e estudados os mecanismos tecnológicos de segurança da informação que poderiam ser relacionados com este fator. Também foram identificados os principais requisitos relacionados à veracidade dos dados e selecionados os que fariam parte da análise e que seriam vinculados aos mecanismos de segurança da informação. Ressalta-se como contribuição deste trabalho a sugestão de inclusão de dois requisitos de veracidade propostos pelos autores deste artigo, além dos requisitos identificados com base na literatura, são eles: a legalidade e a privacidade.

Conforme apresentado, pode-se concluir que os mecanismos: criptografia, controle de acesso, *hashing, backup*, replicação de dados, certificado digital, assinatura digital e carimbo de tempo podem contribuir para os requisitos da veracidade dos dados em ambientes de *Big Data.* Sendo que, para os requisitos de ambiguidade e modelos de aproximação não foram identificados mecanismos de segurança da informação diretamente associados. Recomenda-se para tratar a ambiguidade, ações no âmbito de pessoas e processos e para os modelos de aproximação, quando possível e aplicável, implementar os mecanismos de segurança apresentados para os demais requisitos da veracidade. Conclui-se também que os mecanismos tecnológicos apresentados neste trabalho não tratam os requisitos da veracidade em sua totalidade, mas podem contribuir para este fim. Para uma solução completa devem ser considerados também os mecanismos de segurança não-tecnológicos, ou seja, aqueles relacionados com pessoas e processos.

Este trabalho não possui a intenção de esgotar os assuntos referentes aos mecanismos de segurança da informação e aos requisitos de veracidade dos dados em ambientes de *Big Data*. Mecanismos como autorizações para mudanças nos dados, avaliação nos métodos utilizados para higienização dos dados, métodos utilizados na coleta dos dados, *expertise* do analista dos dados e confiança no fornecedor que armazena os dados são alguns dos pontos a serem considerados para um tratamento mais completo da veracidade dos dados. Como trabalhos futuros sugere-se estudos que possam identificar o quanto a veracidade, sob o viés da segurança da informação, está implementada em ferramentas e processos de *Big Data*, ou ainda, ampliar o estudo incluindo proposições de mecanismos de segurança da informação não tecnológicos que possam contribuir com a veracidade.

## Referências

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS NBR. *ISO/IEC 2002:2013*: Tecnologia da Informação – Técnicas de segurança – Código

de prática para controle de segurança da informação. Rio de Janeiro, 2013.

BAŠKARADAA, Saša; KORONIOSA, Andy. A Critical Success Factor Framework for Information Quality Management. *Information Systems Management***,** v. 31, n. 4, 2014, p. 276-295. Disponível em: https://www.researchgate.net/publication/268335323_A_Critical_Success_Factor_Framework_for_Information_Quality_Management. Acesso em: 20 jun. 2017.

BEAL, Adriana. *Segurança da informação:* princípios e melhores práticas para a proteção dos ativos de informação nas organizações. São Paulo: Atlas, 2005.

CLAVERIE-BERGE, Isabelle. *Solutions Big Data IBM.* 2012. Disponível em http://www-05.ibm.com/fr/events/netezzaDM_2012/Solutions_Big_Data.pdf. Acesso em: 5 ago. 2017.

DAVENPORT, Thomas H. *Big Data @ Work***:** Dispelling the Mhyts, Uncovering the Opportunities. Boston, Massachusetts: Harvard Business School Publishing Corporation, 2014.

DEMÉTRIO, Denise B. *Infra-estrutura de protocolação digital de documentos eletrônicos.* 2003. 142f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Santa Catarina, Florianópolis, 2003. Disponível em: https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/86511/201461.pdf?sequence=1&isAllowed=y. Acesso em: 20 jul. 2017.

DEVAN, Ashley. The 7 V's of *Big Data*. *Impact radius.* 7 abr. 2016. Disponível em: https://www.impactradius.com/blog/7-vs-big-data/. Acesso em: 15 jul. 2017.

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

*FAGUNDES, Priscila B.; MACEDO, Douglas D. J.; FREUND Gislaine P.* A produção científica sobre qualidade de dados em big data: um estudo na base de dados Web of Science*. RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação***,** v. 16, n. 1, 2018.

FERREIRA, Fernando N. F. *Segurança da informação*. Rio de Janeiro: Ciência Moderna, 2003.

FONSECA, João. J. S. *Metodologia da pesquisa científica*. Fortaleza: UEC, 2002.

FONTES, Edison. *Políticas e normas para a segurança da informação:* como desenvolver, implantar e manter regulamentos para a proteção da informação nas organizações. Rio de Janeiro: Brasport, 2012.

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.

GERHARDT, Tatiana E.; SILVEIRA, Denise T. *Métodos de pesquisa***.** Porto Alegre: Editora da UFRGS, 2009.

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

LANEY, Doug. Application Delivery Strategies*. META Group*, 2001. Disponível em: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf. Acesso em: 7 jul. 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.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. *Big data***:** como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana. Rio de Janeiro: Elsevier, 2013.

MCAFEE, Andrew; BRYNJOLFSSON, Erik. Big Data. The management revolution. *Harvard Buiness 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.

MORAES, Alexandre F. *Segurança em redes*: fundamentos. São Paulo: Érica, 2010.

NAKAMURA, Emilio Tissato; GEUS, Paulo Lício de. *Segurança de redes em ambientes cooperativos.* São Paulo: Novatec, 2007.

NETO, Abilio B. O. Desafios de segurança e privacidade em *Big Data*. *IBM developerWorks.* 13 mar. 2016. Disponível em: https://www.ibm.com/developerworks/community/blogs/tlcbr/entry/mp256?lang=en. Acesso em: 11 jul. 2017.

SARACEVIC, T. *Ciência da informação***:** origem, evolução e relações. *Perspectivas em Ciência da Informação*, Belo Horizonte, v.1, n.1, p.41-62, jan./jun.1996. Disponível em: http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22. Acesso em: 16 abr. 2017.

SÊMOLA, Marcos. *Gestão da segurança da informação***:** uma visão executiva. 2. ed. Rio de Janeiro: Elsevier, 2003.

SINGH, Simon. *O livro dos códigos*. 5. ed. Rio de Janeiro: Record, 2005.

SORDI, José Osvaldo de.; MIRELES, Manuel; GRIJO, Rogério Nahas. Gestão da qualidade da informação no contexto das organizações: percepções a partir do experimento de análise da confiabilidade dos

jornais eletrônicos. *Perspectivas em Ciência da Informação*, v. 13, n. 2, 2008, p 168-195. Disponível em: http://www.scielo.br/pdf/pci/v13n2/a12v13n2.pdf. Acesso em: 16 abr. 2017.

WALKER, Michael**.** *Data Veracity*. 2012. Disponível em: http://www.datasciencecentral.com/profiles/blogs/data-veracity. Acesso em 14/07/2017. Acesso em: 11 jul. 2017.

ZIKOPOULOS, Paul. *et al. Understanding Big Data:* analytics for enterprise class hadoop and streaming Data. New York: McGraw-Hill, 2012.