RICI Revista Ibero-Americana de Ciência da Informação

ISSN 1983-5213

Proveniência de dados e segurança da informação: relações interdisciplinares no domínio da Ciência da Informação

Gislaine Parra Freund
Universidade Federal de Santa Catarina, Departamento de Ciência da Informação, Florianópolis, SC, Brasil
gislaineparraf@gmail.com

Márcio José Sembay
Universidade Federal de Santa Catarina, Departamento de Ciência da Informação, Florianópolis, SC, Brasil
marcio.sembay@posgrad.ufsc.br

Douglas Dyllon Jeronimo De Macedo
Universidade Federal de Santa Catarina, Departamento de Ciência da Informação, Florianópolis, SC, Brasil
douglas.macedo@ufsc.br

DOI: https://doi.org/10.26512/rici.v12.n3.2019.21203

Recebido/Recibido/Received: 2018-12-30
Aceitado/Aceptado/Accepted: 2019-05-22

ARTIGOS

Resumo: A proveniência de dados pode ser entendida como o registro que descreve pessoas, instituições, entidades e atividades envolvidas na produção, influência ou entrega de um dado ou objeto. Já a segurança da informação destina-se a preservar o conjunto de informações que representam valor para indivíduos, organizações ou entidades. Nessa perspectiva, esta pesquisa analisa e identifica as relações interdisciplinares existentes entre proveniência de dados e a segurança da informação no domínio da Ciência da Informação. Trata-se de uma pesquisa de natureza básica, caracterizada bibliográfica de caráter exploratório e de abordagem qualitativa. Com o estudo realizado observa-se que a proveniência de dados por si não garante a confidencialidade, a integridade e a disponibilidade necessária no decorrer de seus processos e que as propriedades da segurança da informação contribuem de forma significativa para isso. Da mesma forma, a proveniência de dados pode ser considerada um requisito importante para estabelecer confiabilidade e prover segurança em sistemas computacionais de informação.
Palavras-Chave: Ciência da Informação. Interdisciplinaridade. Proveniência de Dados. Segurança da Informação.

Data Provenance and Security of Information: interdisciplinary relations in the field of Information Science

Abstract: The data provenance can be understood as the record describing people, institutions, entities, and activities involved in the production, influence, or delivery of a data or object. Information security is intended to preserve the set of information that represents value for individuals, organizations or entities. From this perspective, this research analyzes and identifies the interdisciplinary relationships between data origin and information security in the field of Information Science. It is a research of a basic nature, characterized by an exploratory and qualitative approach. With the study carried out it is observed that the provenance of data by itself does not guarantee the confidentiality, the integrity and the necessary availability in the course of its processes and that the properties of the information security contribute in

a significant way for this. Likewise, the provenance of data can be considered an important requirement to establish reliability and provide security in computer information systems.
Keywords: Information Science. Interdisciplinarity. Data Provenance. Information Security.

Procedencia de datos y seguridad de la información: relaciones interdisciplinarias en el ámbito de la ciencia de la información

Resumen: La procedencia de datos puede ser entendida como el registro que describe personas, instituciones, entidades y actividades involucradas en la producción, influencia o entrega de un dato u objeto. La seguridad de la información está destinada a preservar el conjunto de informaciones que representan valor para individuos, organizaciones o entidades. En esta perspectiva, esta investigación analiza e identifica las relaciones interdisciplinarias existentes entre la procedencia de datos y la seguridad de la información en el ámbito de la Ciencia de la Información. Se trata de una investigación de naturaleza básica, caracterizada bibliográfica de carácter exploratorio y de abordaje cualitativo. Con el estudio realizado se observa que la procedencia de datos por sí no garantiza la confidencialidad, la integridad y la disponibilidad necesaria en el transcurso de sus procesos y que las propiedades de la seguridad de la información contribuyen de forma significativa para ello. De la misma forma, la procedencia de datos puede ser considerada un requisito importante para establecer confiabilidad y proveer seguridad en sistemas informáticos de información.
Palabras clave: Ciencia de la Información. Interdisciplinariedad. Procedencia de Datos. Seguridad de la Información.

1 Introdução

O termo proveniência foi inicialmente utilizado no contexto da arte, para descrever a história documentada ou a cadeia de custódia de uma obra de arte, da sua criação até o seu mais recente proprietário (MOREAU *et al*., 2008; TAN, 2008). A proveniência permite que os especialistas possam comprovar a autenticidade de um objeto e influencia diretamente no seu valor (MOREAU *et al*., 2007).

A documentação histórica de um objeto ou dado gerado com a proveniência de dados possibilita que o mesmo obtenha autoridade, permitindo assim que estudiosos compreendam e sejam capazes de avaliar com maior precisão, a importância e o contexto de aplicação daquele objeto e garantir a qualidade e a veracidade dos dados (MOREAU *et al*., 2007; SIMMHAN *et al*., 2005).

Já o tema segurança de informação também está associado a diversas áreas do conhecimento, sua abrangência extrapola os limites dos controles computacionais e seu principal objetivo é proteger as informações em diferentes aspectos. Segundo a norma ABNT NBR ISO/IEC 27002:2013[1], 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*. Em seus controles, a norma apresenta diretrizes destinadas a segurança em sistemas, em redes, em recursos humanos ou seja, apresenta uma abordagem completa contemplando os atores que manipulam a informação.

[1]– Código de prática para controle de segurança da informação.

Ferreira (2003) com uma perspectiva mais corporativa considera que a segurança da informação possibilita a utilização dos recursos que suportam as informações necessárias para as atividades estratégicas, táticas e operacionais de uma organização, de maneira confiável. Já Moraes (2010) complementa esta afirmação com a definição que segurança da informação é um processo para proteger informações do mau uso intencional ou não, realizados por pessoas internas ou externas à organização.

A era digital promoveu reflexões e mudanças nas práticas de segurança da informação tornando o tema relevante e aplicável à inúmeros cenários. Davenport (1998) apresenta que o desafio de prover, usar e disponibilizar dados e informações de forma confiável permeiam os diferentes segmentos da sociedade, seja como provedor de informação, ou como usuário destas.

As informações desempenham um papel importante para o desenvolvimento global e com a ascensão da tecnologia, o volume de dados e informações digitais cresce exponencialmente, fato mencionado por Mayer (2013) como a “avalanche de informação cotidiana" ocasionada pelo uso massivo de dispositivos tecnológicos. Diante disso, acompanhar esse cenário e preocupar-se com a proteção e com o uso seguro das informações é cada vez mais imprescindível para ciência da informação enquanto campo interdisciplinar destinada a investigar os recursos de informação.

Diante do exposto, emerge a seguinte questão de pesquisa: No contexto interdisciplinar existente entre os temas, em quais aspectos a segurança da informação se relaciona com a proveniência de dados no domínio da Ciência da Informação?

Para responder a esta questão, o objetivo geral deste artigo é analisar e identificar os aspectos interdisciplinares existentes entre os tipos, níveis e benefícios da proveniência de dados com as propriedades de segurança da informação, contribuindo para o debate das relações aqui propostas no âmbito da Ciência da Informação, promovendo uma nova visão temática a ser explorada nessa ciência.

Neste estudo serão abordadas as relações entre os temas proveniência de “dados” e segurança da “informação” sendo pertinente ressaltar que diversas abordagens são utilizadas por diferentes autores para definir dados e informação. Para Davenport (1998) dado é a matéria-prima para a informação e informação pode ser compreendida como o dado com significado, ou seja, dotado de relevância e propósito. Porém, neste artigo, “dados” e “informações" fazem parte do escopo de estudo e serão abordados de forma única considerando que ambos precisam ser protegidos e precisos para serem confiáveis e úteis nos mais diversos contextos.

Apesar da proveniência de dados ser aplicável em diferentes cenários desde suas origens, este trabalho aborda o tema e seus processos no contexto de dados digitais e avalia as relações existentes com as abordagens da segurança da informação.

Este artigo está dividido em sete seções, começando por esta Introdução; na segunda seção apresenta a metodologia utilizada; na seção três é apresentada a relação interdisciplinar existente entre os temas Ciência da Informação, Proveniência de Dados e Segurança da Informação; a quarta seção expõe as definições, componentes e aplicações relacionadas à proveniência de dados considerada importante para esse documento; na quinta seção apresenta-se as definições sobre segurança da informação que contribuem para fundamentação desse estudo; a sexta seção é dedicada aos resultados e discussões; a sétima seção apresenta as considerações finais e por último, as referências bibliográficas.

2 Metodologia

Este estudo trata-se de uma pesquisa básica quanto à natureza, pois objetiva gerar conhecimentos novos, úteis para o avanço da ciência, sem aplicação prática prevista (GERHARDT e SILVEIRA, 2009) e como bibliográfica quanto aos procedimentos, que na definição de Fonseca (2002) "é feita a partir do levantamento de referências teóricas já analisadas, e publicadas por meios escritos e eletrônicos, como livros, artigos científicos, páginas de web sites".

Quanto aos objetivos, é de caráter exploratório pois visa obter um melhor entendimento do problema a ser estudado e promover maior familiaridade com os temas, para torná-los mais explícitos ou construir hipóteses (GIL, 1991). Possui uma abordagem qualitativa, visto que se preocupa com o aprofundamento da compreensão de um grupo social, de uma organização, entre outras sem considerar a representatividade numérica (GERHARDT E SILVEIRA, 2009).

Para a identificação dos estudos relacionados com este aqui desenvolvido e a fim de verificar a relevância da pesquisa e a sua contribuição para o meio acadêmico no âmbito da Ciência da Informação, em novembro de 2018 foram realizadas buscas nas bases de dados Brapci, *Library & Information Science Abstracts* (LISA), Scopus e Web of Science, utilizando os descritores "*information security", "data security" e "data protection" vinculado ao descritor* "*data provenance*", no campo de pesquisa tópico o qual contempla a consulta ao título, resumo e as palavras-chave, não limitando o período de tempo das publicações. Os resultados das pesquisas não apresentaram nenhum trabalho publicado no campo da Ciência da Informação. Diante deste resultado, observa-se que os temas ainda não foram explorados nesta área e conforme as relações interdisciplinares no que tange aos propósitos dos temas abordados e a

Ciência da Informação, apresentadas na sessão dois deste artigo, entende-se que o estudo é pertinente e pode contribuir para o avanço de pesquisas sobre os temas.

3 Interdisciplinaridade entre Ciência da Informação e Ciência da Computação

Desde a origem da Ciência da Informação, muitos processos da Ciência da Computação foram incorporados por essa ciência em seus respectivos tempos. Direta ou indiretamente, as tecnologias de informação juntamente com as definições de proveniência foram utilizadas para o desenvolvimento de pesquisas na descrição das origens dos dados, contribuindo assim para o desenvolvimento tanto da área de Ciência da Informação como da Ciência da Computação.

Segundo Borko (1968), a Ciência da Informação se refere ao corpo de conhecimentos relacionados à origem, coleção, organização, armazenamento, recuperação, interpretação, transmissão, transformação e reutilização da informação. Nesse sentido, a proveniência de dados também pactua dessas preocupações dentro de um *workflow* científico, que segundo Deelman *et al.,* (2009) é caracterizado na observação de fenômenos, através da análise de dados, e o uso dos resultados obtidos para provar ou refutar uma hipótese.

Desta forma, vale ressaltar que a preocupação com a origem dos dados e informações é uma constante na Ciência da Informação e também é contemplada nos estudos de Ciência da Computação no contexto da proveniência dos dados, apontando assim mais um quesito da interdisciplinaridade entre as ciências por compartilharem as mesmas preocupações, porém de diferentes pontos de vista e com diferentes aparatos científicos e tecnológicos.

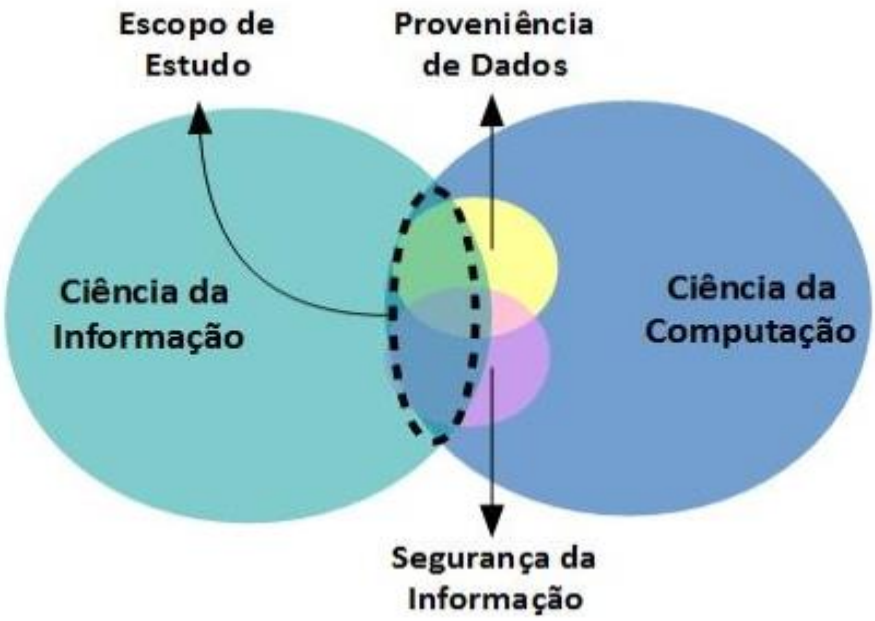
Conforme aponta Alvarenga (2001, p. 12), “mudam-se os meios, sofisticam-se os instrumentos e técnicas e surgem nomes novos para designar coisas que já se faziam no passado”. Entretanto, a essência das coisas permanece. Desse modo, pode-se observar que nos princípios de evolução da tecnologia, o uso de proveniência de dados não é um termo novo, porém pode estar sendo contemplado e pensado com outro contexto.

Este cenário também se aplica à segurança da informação que teve sua origem no contexto das guerras no século XX, pois desde essa época existiam informações que precisavam ser confidenciais e íntegras para a tomada de decisão durante os combates. Esta disciplina foi fortemente empenhada na Ciência da Computação com o desenvolvimento de mecanismos tecnológicos e vem sendo empregada na Ciência da Informação, com mais ênfase a partir da era digital.

Em todos os contextos e áreas, a relevância, a precisão e a integridade das informações são requisitos primordiais. Moraes (2010), ao conceituar segurança da informação, afirma que “toda e qualquer informação deve ser correta, precisa e estar disponível, a fim de ser armazenada, recuperada, manipulada ou processada, além de poder ser trocada de forma

segura e confiável". Esta afirmação combinada com o apontamento do Borko (1968) ao aludir sobre a origem da Ciência da Informação e as premissas da proveniência de dados demonstram a relação interdisciplinar e colaborativa entre os temas e a Ciência no sentido de compartilharem as mesmas preocupações e estarem em busca da confiabilidade dos dados em todo o ciclo de vida. A relação entre as ciências está representada e pode ser visualizada na Figura 1.

Figura 1: Relação entre as ciências



Fonte: Os autores

A área pontilhada da Figura 1 representa a interdisciplinaridade entre a Ciência da Informação, a Ciência da Computação e os temas de proveniência de dados e segurança da informação, sendo este o objeto de exploração que aqui se propõe a analisar no âmbito da Ciência da Informação.

4 Proveniência de Dados

A proveniência de dados possui várias definições, porém, neste estudo foram selecionadas aquelas que mais se aproximam e possibilitam o alinhamento com os aspectos da Ciência da Informação.

O termo proveniência de dados diz respeito à origem ou procedência dos dados e segundo Davidson e Freire (2008) também pode ser referenciado com a auditoria, a trilha, a

linhagem e o pedigree dos dados já que contém as informações do processo que originou o produto destes dados.

Segundo Buneman *et al*.(2001) indica que a proveniência de dados também chamada de linhagem, genealogia ou *pedigree*, consiste em metadados que descrevem as origens de um item de dado, ou seja, referencia o histórico de como aquele dado foi produzido ou derivado. O autor ainda complementa que, a proveniência de dados é a documentação complementar de um determinado dado que contém a descrição de "como", "quando", "onde" e "por que" ele foi obtido e "quem" o obteve.

Nessa mesma linha estão Woodruff *et al*. (1997), afirmando que proveniência não somente inclui a origem do dado (identificação, responsável pelo dado, data de criação), mas também os processos aplicados a ele (algoritmos e seus respectivos parâmetros).

Belhajjame *et al*. (2013) afirma que a proveniência é o registro que descreve pessoas, instituições, entidades e atividades envolvidas na produção, influência ou entrega de um dado ou objeto.

Outro fator importante da proveniência de dados, de acordo com Davidson e Freire (2008), é que a mesma pode ser dividida em três tipos:

- **Prospectiva:** trata-se da sequência de processos utilizados (receita) para a geração do dado, ou seja, captura os passos que devem ser seguidos para a geração de um dado produto.
- **Retrospectiva:** trata-se das informações obtidas durante a execução dos processos de geração do dado. Compreende desde o tempo de duração de cada atividade executada até a origem dos dados de entrada. Além disso, não depende do tratamento da proveniência prospectiva para ser utilizado. Em outras palavras, é como se fosse um *log* detalhado da execução de uma tarefa.
- **Dados definidos pelo usuário:** qualquer informação que o usuário julgar necessária para futura utilização. Como exemplo, pode-se citar anotações, conclusões a respeito do processo e, até mesmo, observações sobre parâmetros utilizados.

Ainda segundo Davidson e Freire (2008), quando a proveniência é capturada de forma automática, pode-se dividi-la nos níveis:

- **Workflow:** envolve a descrição da execução de um processo, ou seja, das tarefas que dele fazem parte, é usado pela grande maioria das soluções com SGWfC (Sistemas de Gerência de Workflow Científicos) e nesse caso deve ser adaptado para capturar os dados dos diferentes processos executados;

- **Atividade:** pode ocorrer de duas formas. Na primeira, cada processo/programa executado é alterado para capturar os dados de proveniência. Na segunda, podem ser criados programas específicos para monitorar a execução de um determinado processo e capturar os dados de proveniência;
- **Sistema Operacional:** utiliza os dados fornecidos pelo próprio sistema operacional como insumo para a proveniência.

E por fim, a proveniência de dados é aplicada nas mais variadas áreas, como nas bibliotecas digitais, na indústria de alimentos, no jornalismo, na rastreabilidade de informações em redes sociais e na transparência de aplicações comerciais, entre outras aplicações (CURBERA *et al*., 2008).

Em experimentos científicos, a proveniência de dados pode ajudar a interpretar e entender resultados: examinando a sequência de passos que levaram ao resultado, dando visibilidade sobre a cadeia de raciocínio utilizada na sua produção, verificar que o experimento foi realizado de acordo com procedimentos aceitáveis, identificar as entradas do experimento e, em alguns casos, reproduzir o resultado (FREIRE *et al*., 2008). Segundo Buneman e Tan (2007), manter um registro completo de como o cálculo ou o processamento foram realizados é essencial para: (a) assegurar a repetitividade, (b) catalogar o resultado, (c) evitar a duplicação de esforços, e (d) recuperar as fontes de dados a partir dos dados de saída.

Segundo Bose e Frew (2005), os principais benefícios da proveniência para a qualidade de dados são:

- **Comunica a qualidade de dados:** confiabilidade, adequação, acurácia, atualidade, redundância;
- **Melhora a interpretação do dado**: em relação a função do reconhecimento da fonte e na utilização do dado para um aspecto de tomada de decisão;
- **Justificativa do uso de um determinado dado:** em relação as limitações e intenções originais do uso de um determinado dado de conjuntos de dados ambientais;
- **Redução de erros:** no quesito do juízo da precisão do dado, no acompanhamento preciso da linhagem de conjuntos de dados científicos;
- **Passos do processamento:** permite que usuários não especialistas em dados entendam a capacidade de recuperar e entender os relacionamentos entre produtos de dados, scripts ou dados gerados por programas;
- **Criação de dados científicos:** permite identificar o processo utilizado para ajudar a identificar e avaliar os componentes básicos dos sistemas que fornecem a recuperação de linhagem para produtos de dados científicos;
- **Atualização de dados:** permite a partir do desenvolvimento de estudos formais para executar rastreamento de linhagem de dados em visões relacionais;
- **Modificação de *schemas* de visões relacionais:** modelos gráficos e estudos experimentais;
- **Fontes de dados históricas:** permite identificar a origem e o subsequente histórico de processamento.

Diante disso, a proveniência de dados vem se tornando cada vez mais presente no ambiente científico, tanto para garantir a origem dos dados como para avaliar a sua acurácia. Importante ressaltar que antes de definir quais informações são necessárias para garantir a proveniência e como serão gerenciadas, é preciso definir uma forma de organizá-las a fim de que se possa, posteriormente, recuperá-las e entendê-las para que tragam o máximo de benefícios possíveis (ALMEIDA, 2015).

Ainda Almeida (2015) ressalta que os modelos de proveniência de dados têm como principal objetivo fornecer uma estrutura para que os dados de proveniência possam ser armazenados e recuperados, mantendo seu significado e potencializando os seus benefícios.

Além da aplicação dos benefícios da proveniência de dados em diversas áreas científicas, existem também, diversos modelos que podem ser aplicados em situações adversas. Podemos citar três modelos dos quais existem várias aplicações conforme encontrados na literatura:

- O modelo W7 proposto por Ram e Liu (2007) tem como base a ontologia a qual objetiva descrever as propriedades de um objeto de caráter geral.
- O modelo *Provenance Vocabulary* apresentado por Hartig e Zhao (2010) volta sua atenção para o problema da proveniência de dados publicados na web.
- O modelo proposto por Sahoo e Sheth (2009) intitulado *Provenir Ontology* foi desenvolvido para ser um modelo de proveniência de dados genético, priorizando a interoperabilidade entre diferentes sistemas e sua adaptação para qualquer aplicação.

Na próxima seção são apresentadas algumas definições sobre segurança da informação que contribuem para fundamentação desse estudo.

5 Segurança da Informação

Na compreensão de Sêmola (2003) segurança da informação é “uma área do conhecimento dedicada à proteção dos ativos de informação contra acessos não autorizados, alterações indevidas ou sua indisponibilidade”.

A segurança da informação possui propriedades que devem ser compreendidas e consideradas em sua implantação. Por questões históricas, o termo segurança da informação remete inicialmente à propriedade da confidencialidade. Desde suas origens, na época das guerras já existiam informações críticas e essenciais para o sucesso das batalhas que demandavam níveis altos de sigilo e que mesmo de forma rudimentar necessitaram de controles de segurança (SINGH, 2005). Desde então, a confidencialidade é uma preocupação eminente e a mais lembrada, relacionada ao tema. Porém, outras foram sendo incluídas e consideradas para a obtenção da segurança em um âmbito global.

Sêmola (2003) as denominam de princípios e apresenta que são três os princípios básicos que norteiam a implementação da segurança da informação: Confidencialidade,

Integridade e Disponibilidade. Na visão de Beal (2005) informações críticas devem ser protegidas para evitar além do acesso, divulgação e adulteração não autorizadas, sua destruição inadequada e indisponibilidade temporária. Fontes (2012) também ressalta que disponibilidade, integridade e confidencialidade protegerá a informação para que a organização operacionalize os seus negócios e atenda a seus objetivos.

Já Nakamura e Geus (2007) apresentam duas propriedades adicionais para que a proteção da informação seja alcançada: a Autenticidade e o Não-repúdio - também encontrada na literatura como Irretratabilidade.

Observando as similaridades de nomenclatura e de interpretação existente entre os autores, as propriedades de segurança da informação podem ser compreendidas como:

- **Confidencialidade:** garante que a informação seja disponibilizada ou divulgada somente a indivíduos, entidades ou processos autorizados.
- **Integridade:** assegura que a informação é íntegra (completa) e fidedigna (exata) ou seja, que ela não sofra alterações por entidade não autorizada pelo seu proprietário.
- **Disponibilidade:** assegura que a informação esteja acessível, disponível e utilizável sempre que for necessária e demandada.
- **Autenticidade:** Garante que a informação é proveniente da fonte anunciada, ou seja não pode sofrer modificações ao longo do processo quanto a sua origem.
- **Irretratabilidade:** também denominada como não repúdio, refere-se à garantia que uma entidade não negue a autoria de algo realizado por ela.

De acordo com Baars *et al.* (2015), a tríade Confidencialidade, Integridade e Disponibilidade da Informação está atribuída ao aspecto da Confiabilidade da Informação e para cada uma das três propriedades, os autores apresentam suas características.

Em relação à propriedade Confidencialidade, apresentam as características da exclusividade e privacidade, sendo:

- **Exclusividade:** dados disponíveis exclusivamente a usuários autorizados a acessá-los; e
- **Privacidade:** restrição de acesso à dados pessoais.

Já relacionado à propriedade de Integridade expõem as características da completeza, corretude, precisão, validade e verificação, sendo:

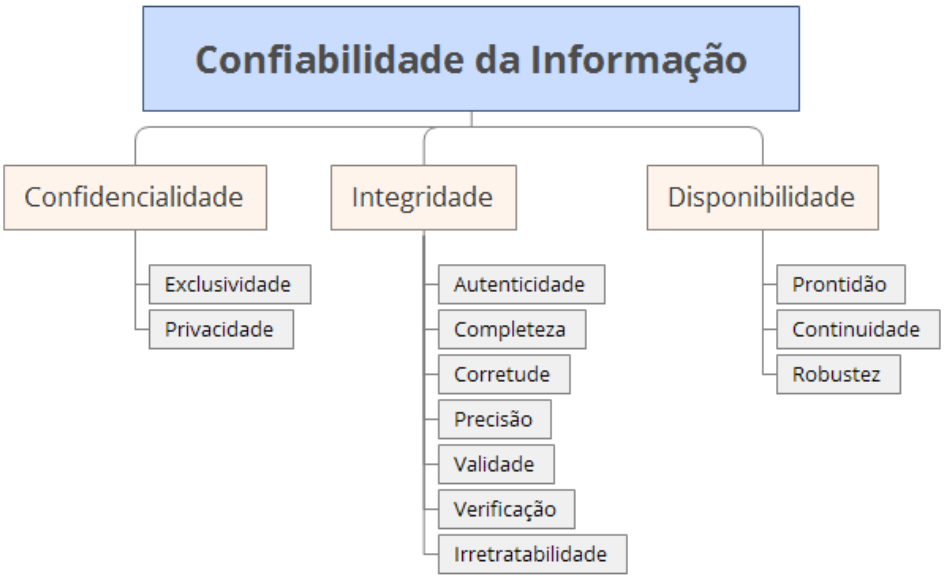
- **Completeza:** dados completos e inteiros.
- **Corretude:** dados exatos e verdadeiros.
- **Precisão:** dados (de saída) consistentes.
- **Validade:** dados compatíveis com os critérios de aceitação (exatidão, precisão, tempo de vida)
- **Verificação:** possibilidade de aferição e rastreabilidade do fluxo dos dados (cadastro, armazenamento, recuperação, transferência e exibição dos dados corretamente).

Referente à propriedade de Disponibilidade, as principais características apresentadas pelos autores são prontidão, continuidade e robustez, sendo:

- **Prontidão:** sistemas disponíveis sempre que forem demandados.
- **Continuidade:** pessoal preparado para continuar seu trabalho mesmo em situações de anomalias.
- **Robustez:** sistemas com capacidade de atender as demandas necessárias.

Baars *et al.* (2015) apresentam a abordagem das propriedades de segurança de forma mais detalhada. Diante disso, os aspectos da autenticidade e irretratabilidade foram incorporados à propriedade da Integridade conforme representada na Figura 2 e esta abordagem será utilizada como referência para a análise que mostrará a influência e relação entre os temas de proveniência de dados e segurança da informação.

Figura 2: Confiabilidade da informação



Fonte: Os autores – 2018.

A Figura 2 apresenta a abordagem integrada das propriedades de segurança da informação e suas características que são utilizadas para nortear as análises deste estudo.

6 Resultados e Discussões

A seguir são apresentadas as relações identificadas entre os tipos, níveis e benefícios da proveniência de dados, e as propriedades da confidencialidade, integridade e disponibilidade referentes à segurança da informação.

Procurou-se estabelecer uma relação entre a proveniência de dados e as propriedades da segurança da informação que de alguma forma fossem complementares, que contribuísse com a confiabilidade da informação já que este é um objetivo comum de ambos os temas.

As relações foram identificadas considerando que as propriedades da segurança da informação podem ser necessárias à proveniência de dados, a todos os tipos de dados, representado pelo símbolo ⊗; pode estar condicionado à legalidade vinculada ao dado e/ou ser necessária conforme requerido pelo tipo de dado (sensível ou não) ou ao cenário de aplicação (considerando que o dado pode não ser considerado sensível mas o processo sim), representado pelo símbolo ○; ou não se aplica ou seja, o aspecto não possui relação, representado pelo símbolo ⊘. Estas informações estão dispostas no quadro em forma de legenda.

O Quadro 1 exibe um resumo das relações identificadas e em seguida são apresentadas as justificativas para cada uma delas, sendo que esse mesmo quadro dispõe nas linhas, os tipos, os níveis e os benefícios da proveniência de dados e nas colunas, denominados como aspectos, as características das propriedades da Confidencialidade, Integridade e Disponibilidade da Informação.

Quadro 1 – Proveniência de Dados, Segurança da Informação e suas relações

PROVENIÊNCIA DE DADOS			SEGURANÇA DA INFORMAÇÃO											
			CONFIDENCIALIDADE		INTEGRIDADE							DISPONIBILIDADE		
			Exclusividade	Privacidade	Autenticidade	Completeza	Corretude	Precisão	Validade	Verificação	Irretratabilidade	Prontidão	Continuidade	Robustez
	TIPOS	Prospectiva	⊗	○	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊘	⊘	⊗
		Retrospectiva	⊗	○	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊘	⊘	⊗
		Dados (input usuários)	⊗	○	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊘	⊘	⊘
	NÍVEIS	Workflow	○	○	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊘	⊗
		Atividade	○	○	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊘	⊗
		Sistema Operacional	○	○	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊘	⊗
	BENEFÍCIOS	Comunicação Qualidade dos Dados	○	○	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊘	⊗
		Reconhecimento da Fonte	○	○	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊘	⊗
		Uso de Dados	○	○	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊘	⊗
		Precisão dos Dados	○	○	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊘	⊗
		Compreensão do Processamento	○	○	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊘	⊗
		Visibilidade do Processo de Criação dos Dados	○	○	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊘	⊗
		Visões Relacionais (atualização dos dados)	○	○	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊘	⊗
		Modificações dos *Schemas*	○	○	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊘	⊗
		Fontes de Dados	○	○	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊘	⊗

⊗ - Necessário a todos os tipos de dados. ○ - Condicionado a legalidade vinculada ao dado e/ou conforme o tipo de dado ou cenário de aplicação ⊘ - Não se aplica.

Fonte: Os autores - 2018

Ao analisar a relação da segurança da informação sob a perspectiva dos tipos de proveniência de dados com a propriedade da confidencialidade, na proveniência prospectiva, retrospectiva e nos dados, o aspecto da exclusividade é requerido independentemente do tipo de dado manipulado e do cenário, pois tanto na construção do percurso (prospectiva) quanto no rastreamento do percurso trilhado pelos dados (retrospectiva) e nos dados inseridos pelo usuário, é importante assegurar que somente entidades autorizadas acessem os dados e incluam, alterem ou excluam. Na proveniência retrospectiva, por se tratar do registro dos eventos de construção de algo, por motivos estratégicos, a necessidade de sigilo pode ser considerada mais evidente. Já o aspecto da privacidade, em todos os tipos (prospectiva, retrospectiva e dados) apresentou-se necessário nos casos em que a proveniência contemple dados pessoais os quais requerem privacidade por força de lei.

Quantos aos tipos de proveniência de dados, com a propriedade da integridade, na proveniência prospectiva, na retrospectiva e dados, todos os aspectos de segurança (autenticidade, completeza, corretude, precisão, validade, verificação e irretratabilidade) apresentaram relação como necessários independente do tipo de dados. Esta forte relação se esclarece ao observar as características da integridade e o propósito da proveniência de dados as quais apresentam grande proximidade de objetivos de forma que possuam relação com todos os tipos de proveniência.

Já ao analisar a relação dos tipos de proveniência de dados com a propriedade da disponibilidade, observa-se que somente o aspecto robustez se relaciona com a proveniência prospectiva e retrospectiva sendo considerados necessários para todos os tipos de dados. Os aspectos da prontidão e continuidade não se aplicam a nenhum dos tipos de proveniência, pois se trata de aspectos operacionais dos sistemas e pessoas que os operam e que não apresentam relação direta com o propósito da proveniência de dados.

No que se refere à perspectiva dos níveis de proveniência de dados com a propriedade da confidencialidade observa-se que tanto nos níveis de *workflow* o qual registra as tarefas de um processo com a captura de dados, no nível da atividade a qual consiste no monitoramento do processo e na captura de dados propriamente dita, quanto no nível de sistema operacional o qual fornece os dados para a proveniência, a propriedade de exclusividade e de privacidade são aspectos condicionados ao tipo dos dados. Em todos os níveis, o controle de acesso aos dados referenciado na propriedade da exclusividade é necessário somente se forem considerados sensíveis e/ou se o cenário de aplicação demande exclusividade de acesso a eles. Da mesma forma a privacidade será necessária somente se o cenário de aplicação manipule dados de caráter pessoal, sendo necessária para atendimento da legislação.

Quanto à relação dos níveis de proveniência de dados com a propriedade da integridade, da mesma forma observada na avaliação quanto aos tipos de proveniência, em todos os níveis, todos os aspectos relacionados à integridade foram considerados necessários, independente do tipo de dados. Para que a proveniência seja válida e confiável é necessário que tanto nos níveis de workflow, da atividade ou do sistema operacional o dado seja autêntico, completo, correto, preciso, válido, passível de verificação e que não possa ser repudiado pela origem.

No que se refere à proveniência de dados e a propriedade da disponibilidade, verifica-se que por se tratar de captura, fornecimento e registro de dados, a prontidão e a robustez dos sistemas apresentam-se como aspectos necessários independente do tipo de dado manipulado. Afinal, para que os propósitos da proveniência sejam executados em cada um dos níveis, os sistemas necessitam estar disponíveis sempre que forem demandados e que possuam recursos necessários para atender a essas demandas. Referente ao aspecto da continuidade, observa-se que não é aplicável aos níveis, por estes tratarem a proveniência somente no âmbito de sistemas.

Já os benefícios da proveniência de dados estão relacionados com a propriedade da confidencialidade, em todos os itens avaliados. A exclusividade e a privacidade são aspectos condicionados a legalidade vinculada ao dado e/ou conforme o tipo de dado ou cenário de aplicação considerando que em alguns cenários os dados manipulados não requerem controle de acesso e sigilo, porém não são todos eles que apresentam essa característica e nem todos os dados estão disponíveis em acesso aberto, podendo em muitos casos possuírem restrições de acesso e uso.

Nessa mesma análise, contemplando a propriedade de integridade, pôde-se observar que a autenticidade, completeza, corretude, precisão, validade, verificação e irretratabilidade, são aspectos necessários a todos os tipos de dados em todos os itens relacionados aos benefícios da proveniência de dados por se tratar de aspectos referentes a qualidade dos dados.

Em relação à propriedade de disponibilidade, verifica-se que para os benefícios da proveniência de dados, a prontidão e a robustez são aspectos necessários a todos os tipos de dados, ou seja, para que a proveniência de dados gere benefícios, os sistemas devem estar disponíveis sempre que forem requisitados e que atendam as demandas necessárias aos processos que geram o workflow dos dados digitais. Já ao que se refere a continuidade não se aplica a nenhum benefício, pois nem sempre teremos pessoas ou equipes especializadas para determinadas situações, sendo sempre necessário realizar adaptações de conhecimento nos mais variados cenários de aplicação.

7 Considerações Finais

Com base no estudo realizado é possível observar que a proveniência de dados pode ser considerada um requisito importante para estabelecer confiabilidade e prover segurança em sistemas computacionais de informação. Por outro lado, a segurança da informação possui propriedades que estão alinhadas e apresentam fortes relações com a proveniência de dados considerando que esta utiliza variados métodos de captura de dados em sistemas de informação, de forma que as propriedades de segurança são importantes e favorecem na qualidade dos dados da proveniência.

A análise realizada neste estudo mostra que a propriedade da integridade com suas características destinadas à completeza e exatidão dos dados, apresentou relação com todos os itens da proveniência, sem restrição quanto ao tipo de dado e cenário, afinal, em todas as situações, para que os dados da proveniência sejam úteis em qualquer contexto, precisam ser confiáveis. A propriedade da confidencialidade também apresentou relação com todos os itens da proveniência que foram avaliados, porém sua necessidade advém do tipo de dados e do cenário de aplicação, podendo estes serem sensíveis ou não, ou se tratar de dados pessoais os quais demandas privacidade por força de lei. Sendo assim, na confidencialidade a maioria das relações identificadas estão condicionadas ao tipo de dado manipulado. Em relação à propriedade da disponibilidade, observa-se que nem sempre todas as características podem ser aplicadas nos tipos, níveis e benefícios da proveniência de dados. Para este caso, a continuidade apresentou-se como uma propriedade que não se aplica aos itens avaliados por se tratar de procedimentos metodológicos que envolvem pessoas e está fora do escopo dos sistemas e dados digitais a qual se propõe este estudo.

Diante do exposto, com base no objetivo aqui proposto, observa-se na análise realizada que as propriedades da segurança da informação têm sua aplicação de forma significativa nos quesitos analisados na proveniência de dados. Verifica-se que a proveniência de dados por si não garante a confidencialidade, a integridade e a disponibilidade necessária no decorrer de seus processos, de forma que a contribuição das propriedades de segurança da informação seja evidente para que o processo da proveniência de dados tenha êxito.

A partir da revisão da literatura realizada, verificou-se que o estudo de proveniência de dados em conjunto com a segurança da informação são temas que ainda não foram explorados no âmbito da Ciência da Informação e que diante das relações existentes entre eles, tratam-se de temáticas importantes a serem investigadas que contribuirão para o avanço desta Ciência.

Como trabalhos futuros, sugere-se que seja realizada uma análise considerando as propriedades de segurança da informação com os modelos de proveniência de dados existentes para

identificar quais delas são adotadas e em qual nível são praticadas. Propõe-se também que seja realizado um estudo que identifique os mecanismos tecnológicos que podem ser adotados para a implementação das propriedades de segurança nos processos da proveniência de dados.

Referências

ALMEIDA, Rodrigo Pinheiro de. **Gerenciamento de Dados de Proveniência de Workflow de Bioinformática com Banco de Dados Baseados em Grafo.** 2015. 139 f. Dissertação (Mestrado) - Curso de Informática. Disponível em: <http://repositorio.unb.br/bitstream/10482/22029/1/2015_RodrigoPinheirodeAlmeida.pdf>. Acesso em: 15 dez. 2018.

ALVARENGA, L. A teoria do conceito revisada em conexão com ontologias e metadados no contexto das Bibliotecas tradicionais e digitais. **Data Grama Zero– Revista de Ciência da Informação**, Rio de Janeiro, v. 2, n. 6, dez. 2001. Disponível em: http://www.brapci.inf.br/index.php/article/download/7457 Acesso em: 4 nov. 2018.

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.

BAARS, Hans; HINTZBERGEN, Jule; SMULDERS, André; HINTZBERGEN, Kees**. Foundations of Information Security Based on ISO 27001 and ISO 27002.** 3rd. ed. Zaltbommel: Van Haren Publishing, 2015.

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

BELHAJJAME, K. *et al.* PROV-DM: The PROV Data Model. **W3C Recommendation**, 30 abr. 2013. Disponível em: https://www.w3.org/TR/prov-dm/. Acesso em: 01 nov. 2018.

BORKO, H. Information Science: What is it? **American Documentation**, v. 19, n. 1, p. 3-5, Jan. 1968.

BOSE, R; FREW, J. 2005. Lineage retrieval for scientific data processing: a survey. **ACM Comput. Surv**. v. 37, n. 1, p. 1-28, March 2005. Disponível em: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=C1746A7A546FD1B64E83540C6ED55688?doi=10.1.1.103.4855&rep=rep1&type=pdf. Acesso em: 29 out. 2018.

BUNEMAN, P., KHANNA, S. E CHIEW, W. (2001). Why and Where: a Characterization of Data Provenance. In: ICDT'01, 8th International Conference on Database Theory, LNCS, v.1973, p. 316–330. Disponível em: https://pdfs.semanticscholar.org/b647/2ee11749ef70713bfb0e322a9ec27523ed88.pdf Acesso em: 01 out. 2018.

BUNEMAN, P.; W.TAN. **Provenance in databases**. In: Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, China, 2007. Disponível em: https://pdfs.semanticscholar.org/87c1/84ab63bf6b03cff9ca2afb385d07e2a8faff.pdf Acesso em: 03 nov. 2018.

CURBERA, F. *et al*. Business Provenance – A Technology to Increase Traceability of End-to-End Operations. In: MEERSMAN, R.; TARI, Z. (Ed.) . **On the Move to Meaningful Internet Systems: OTM 2008.** Lecture Notes in Computer Science. [s.l.] Berlin, Heidelberg: Springer, 2008. p. 100–119. Disponível em: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.473.2548&rep=rep1&type=pdf Acesso em: 15 nov. 2018.

DAVENPORT, T. H. **Ecologia da informação:** porque só a tecnologia não basta para o sucesso na era da informação. São Paulo: Futura, 1998.

DAVIDSON, S. B; FREIRE, J. (2008). **Provenance and scientific workflows: challenges and opportunities**. In Proceedings of the 2008 ACM SIGMOD international conference on Management of data. p. 1345–1350. Disponível em: https://vgc.poly.edu/~juliana/pub/freire-tutorial-sigmod2008.pdf Acesso em: 02 set. 2018.

DEELMAN, E., GANNON, D., SHIELDS, M., TAYLOR, I. (2009), Workflows and e-Science: An overview of workflow system features and capabilities. **Future Generation Computer Systems**, v. 25, n. 5, p. 528–540. Disponível em: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.165.9173&rep=rep1&type=pdf Acesso em: 03 nov. 2018.

FERREIRA, Fernando N. Freitas. **Segurança da informação**. Rio de Janeiro: Ciência Moderna, c2003.
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.

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

FREIRE, J.; KOOP, D.; SANTOS, E.; SILVA, C. T. Provenance for Computational Tasks: A Survey. **Journal Computing in Science and Engineering,** v. 10, n. 3, p. 11-21, 2008. Disponível em: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.147.3801&rep=rep1&type=pdf Acesso em: 24 out. 2018.

GERHARDT, TATIANA ENGEL; SILVEIRA, DENISE TOLFO. **Métodos de pesquisa.** Porto Alegre: Editora da UFRGS, 2009.

GIL, ANTONIO CARLOS. **Como elaborar projetos de pesquisa**. 3. ed. São Paulo: Atlas, 1991.

HARTIG, O; ZHAO, J. Publishing and consuming provenance metadata on the web of linked data. In: **Provenance and annotation of data and processes.** P. 78–90. Springer, 2010. Disponível em: https://link.springer.com/content/pdf/10.1007%2F978-3-642-17819-1_10.pdf. Acesso em: 23 nov. 2018

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

MORAES, Alexandre Fernandes de. **Segurança em redes: fundamentos**. São Paulo: Érica, c2010

MOREAU L, FREIRE J, FUTRELLE J, MCGRATH RE, MYERS J, PAULSON P. **The Open Provenance Model:** An Overview. IPAW, LNCS 5272, , 2008, Berlin, Heidelberg: Springer, 2008. p. 323–326 Disponível em: https://pdfs.semanticscholar.org/27ea/5b41d90da547563db4ceaaddcba9c6b9a506.pdf Acesso em: 29 set. 2018.

MOREAU L, GROTH P, MILES S, VAZQUEZ-SALCEDA J, IBBOTSON J, JIANG S, MUNROE S, RANA O, SCHREIBER A, TAN V, VARGA L**. The provenance of electronic data.** Communications of the ACM 2007, v. 4, p. 52-58, 2007. Disponível em: https://cow.ceng.metu.edu.tr/Courses/download_courseFile.php?id=6950 Acesso em: 30 out. 2018.
NAKAMURA, Emilio Tissato; GEUS, Paulo Lício de. **Segurança de redes em ambientes cooperativos.** São Paulo: Novatec, 2007.

RAM, S; LIU, J. Understanding the semantics of data provenance to support active conceptual modeling. In**: Active conceptual modeling of learning**, pag 17–29, 2007. Springer. Disponível em: https://pdfs.semanticscholar.org/5cb2/e551f7cfe52c13e8bafc0f2e5ebbb65c010f.pdf Acesso em: 02 out. 2018.

SAHOO, S. S; SHETH, A. P. **Provenir ontology: Towards a framework for science provenance management.** Kno.e.sis Publications, 2009. Disponível em: https://corescholar.libraries.wright.edu/cgi/viewcontent.cgi?article=1079&context=knoesis Acesso em: 18 nov. 2018.

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

SIMMHAN YL, PLALE B, GANNON D. **A Survey of Data Provenance Techniques**. In: Technical Report TR-618: Computer Science Department; Indiana University, 2005. Disponível em: ftp://ftp.extreme.indiana.edu/pub/techreports/TR618.pdf Acesso em: 04 dez. 2018.

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

WOODRUFF A, STONEBRAKER M: **Supporting Fine-Grained Data Lineage in a Database Visualization.** Em International Conference on Data Engineering; Birmingham, UK. 1997:15, april 1997. Disponível em: http://db.cs.berkeley.edu/papers/CSD-97-932.pdf. Acesso em: 24 nov. 2018.