

UNIVERSIDADE FEDERAL DO PARANÁ

LUIZ HENRIQUE KRENISKI GUNTZEL

GOVERNANÇA DE DADOS E O PAPEL DO *DATA OWNER*: UMA REVISÃO
INTEGRATIVA

CURITIBA

2023

LUIZ HENRIQUE KRENISKI GUNTZEL

GOVERNANÇA DE DADOS E O PAPEL DO *DATA OWNER*: UMA REVISÃO
INTEGRATIVA

Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção de grau de Bacharel em Gestão da Informação no curso de graduação em Gestão da Informação do Departamento de Ciência e Gestão da Informação do Setor de Ciências Sociais Aplicadas da Universidade Federal do Paraná.

Orientadora: Prof^a. Dr^a. Taiane Ritta Coelho.

CURITIBA

2023

AGRADECIMENTOS

Agradeço à Universidade Federal do Paraná, ao curso de Gestão da Informação e a todos os seus docentes pela oportunidade de uma formação acadêmica de qualidade.

Agradeço à professora Dra. Taiane Ritta Coelho por seus ensinamentos, paciência e orientação tanto durante a realização dessa pesquisa, como também por sua excelente tutoria desde o início da minha trajetória no curso.

Agradeço à minha família pelo suporte desde sempre, pela compreensão e paciência, pela valorização da educação, por todos seus ensinamentos e motivação, me guiando sempre para o melhor caminho.

Agradeço à minha namorada, por estar ao meu lado durante toda a minha trajetória acadêmica, pelo seu suporte e dedicação, me incentivando e apoiando na realização dessa pesquisa.

Agradeço aos meus colegas e amigos, pelo apoio para superar os desafios durante toda minha trajetória acadêmica.

RESUMO

Estudo sobre o papel do *data owner* para a governança de dados nas organizações, buscando identificar a necessidade do papel, a evolução desse dentro da perspectiva da governança de dados e o exercício desse papel na governança de dados das organizações. Realiza revisão integrativa da literatura juntamente com conferências e livros atuais com o objetivo de reunir os diversos ângulos apresentados sobre o papel do *data owner* na governança de dados na literatura e prática para esclarecer as melhores práticas existentes tanto na determinação do papel como suas responsabilidades. Esclarece a evolução nas últimas décadas, da forma como organizações gerenciam e utilizam seus dados, assim como recente transformação significativa, a partir do crescimento exponencial dos dados juntamente com a utilização de tecnologias de ponta, que tornaram abordagens tradicionais de gestão e governança de dados insuficientes, principalmente com a explosão da internet, computação e nuvem, big data e analytics, e outras novas tecnologias, deslocando o prisma da governança de problemas comuns de armazenamento e recuperação, para complexos obstáculos na interoperabilidade, integridade, qualidade, acurácia e usabilidade dos dados. Elevando a importância de uma efetiva gestão e governança de dados com papéis e responsabilidades. Pesquisa revela novos modelos emergentes para gestão e governança de dados em alto volume e velocidade, que alinham a estratégia organizacional, políticas, com posições de autoridade e controle promovendo uma melhor tomada de decisão em ambientes mais complexos e desafiadores.

Palavras-chave: Gestão de dados. Governança de dados. *Big Data Analytics*. *Data Owner*. *Data Mesh*.

ABSTRACT

This study investigates the role of the data owner to the data governance in organizations, aims to identify the necessity of the role, evolution of this role inside the data governance perspective and the exercise of this role in data governance on organizations. Conducts a study of integrative literature review along with the analysis of conferences and current books with the objective of gathering the different existent angles of the role of data owner in the literature and practice to clarify the best practices and clarify their role and responsibilities. Clarifies the evolution within the last decades, in the form of how organizations manage and use their data, also with recent significant transformations, from exponential growth of data along with use of high end technologies, that made traditional approaches of data management and governance insufficient, with the explosion of the internet, cloud computing, big data and analytics and other new technologies, shifting the prism of the governance from usual issues of storage and recovery, to more complex obstacles in the interoperability, integrity, quality, accuracy and usability of data. Elevating the importance of an effective management and governance of data with the correct roles and responsibilities. Research uncovers new emergent models for data management and governance for high volume and velocity data, that aligns with business strategy, policies, authoritative positions and controls promoting a better decision-making in complex and challenging environments.

Keywords: Data Management. Data Governance. Big Data Analytics. Data Owner. Data Mesh.

LISTA DE FIGURAS

FIGURA 1 - PUBLICAÇÕES RELEVANTES EM GOVERNANÇA DE DADOS	12
FIGURA 2 - BARREIRAS PARA IMPLEMENTAÇÃO DA GOVERNANÇA	14
FIGURA 3 - PROPOSTA DE CICLO DE VIDA DE <i>BIG DATA ANALYTICS</i>	20
FIGURA 4 - CICLO DE VIDA PROPOSTO <i>SMART DATA LIFECYCLE</i>	21
FIGURA 5 - CONCEITOS RELATIVOS A MODELOS DE GOVERNANÇA.....	24
FIGURA 6 - ESTRUTURAS DE GOVERNANÇA DE DADOS	25
FIGURA 7 - CARACTERIZAÇÃO DA PESQUISA	28
FIGURA 8 - ARTIGOS POR ÁREA DE CONHECIMENTO.....	32
FIGURA 9 - ARTIGOS POR PAÍS (SCIENCEDIRECT)	33
FIGURA 10 - ORGANIZAÇÃO ESTRUTURAL	40
FIGURA 11 - MUDANÇAS ORGANIZACIONAIS DATA MESH.....	43
FIGURA 12 - INTERCONEXÃO PRINCÍPIOS DATA MESH.....	45

LISTA DE QUADROS

TABELA 1 – BUSCA E COLETA DE ARTIGOS NA LITERATURA.....	29
TABELA 2 – PRINCÍPIOS DO MODELO <i>DATA MESH</i>	42

SUMÁRIO

1 INTRODUÇÃO	9
1.1 OBJETIVOS	10
1.1.1 Objetivo geral	10
1.1.2 Objetivos específicos.....	11
1.2 JUSTIFICATIVA	11
1.2.1 Aspecto Científico	11
1.2.2 Aspecto Socioeconômico	13
2 REFERENCIAL TEÓRICO	16
2.1 GESTÃO DE DADOS	16
2.1.1 Dados como Ativo Organizacional	17
2.1.2 Ciclo de Vida dos Dados	18
2.2 GOVERNANÇA DE DADOS	22
2.2.1 Modelos e Estrutura de Governança	23
2.2.2 Data Stewards e <i>Data Owner</i>	25
3 METODOLOGIA	27
3.1 CARACTERIZAÇÃO DA PESQUISA	27
3.2 MÉTODO E PROCEDIMENTOS	28
3.3 BUSCA NA LITERATURA.....	29
3.4 COLETA DE INFORMAÇÕES DE CONFERÊNCIA E OUTROS	30
4 RESULTADOS	32
4.1 ANÁLISE DOS RESULTADOS: EVOLUÇÃO DA LITERATURA	32
4.1.1 TENDÊNCIAS EM GOVERNANÇA DESCENTRALIZADA: SAÚDE E IOT	33
4.1.2 GOVERNANÇA E BLOCKCHAIN	36
4.2 ANÁLISE DOS RESULTADOS: CONFERÊNCIA E TENDÊNCIAS DA ÁREA ...	37
4.2.1 CONFERÊNCIA ENTERPRISE DATA WORLD.....	37
4.2.2 GOVERNANÇA FEDERADA	41
4.2.3 ANÁLISE SOBRE A GOVERNANÇA FEDERADA: DATA MESH.....	42
4.3 DISCUSSÃO DOS RESULTADOS	47
5 CONSIDERAÇÕES FINAIS	49
REFERÊNCIAS	51

1 INTRODUÇÃO

Dados hoje são cada vez mais tratados como ativos organizacionais, essa transformação cria um novo paradigma de desafios para criação, armazenamento, controle, distribuição e uso dos dados dentro das organizações, existindo assim a emergente necessidade da gestão desses dados assim como demais ativos estratégicos das organizações para proteção e contínua obtenção de valor (EVANS; PRICE, 2012).

Organizações dependem cada vez mais de seus dados para uma melhor tomada de decisão, eficiência e eficácia operacional, para entendimento aprofundado sobre seus negócios, criação de novos produtos e serviços, controle de riscos, dentre outros (REDMAN, 2008). Sendo a realidade de empresas, instituições governamentais e até mesmo para o terceiro setor. Portanto, tendo a tomada de decisão movidas pelos seus dados, existe também a inerente necessidade da preocupação com toda a estrutura necessária para a gestão adequada desses dados (AIKEN; HARBOUR, 2017).

A gestão de dados não ocorre de forma acidental (DAMA, 2017), existe a inerente necessidade de liderança e gestão de políticas e práticas, o exercício de autoridade e controle, para garantir e guiar as boas práticas que sustentam de forma contínua e segura o valor dos dados nas organizações. (LADLEY, 2012).

Ao se tratar de governança de dados, seu exercício exige autoridade e controle para o planejamento, monitoramento e execução da gestão de dados. A governança não tem como objetivo a si própria, e, portanto, advém de uma estratégia organizacional sólida para existir, assim como de um alinhamento estratégico para perdurar. As organizações tomam decisões sobre seus dados independente de possuírem formalidades, porém o exercício formal de autoridade e controle possui maior intencionalidade, credibilidade e efetividade (SEINER, 2014).

Os mecanismos estruturais de uma governança de dados necessitam da definição de papéis e responsabilidades organizados e estruturados (BORGMAN et al., 2016) permitindo a formação de programas de estratégia, políticas e normas, *compliance* e auditorias, identificação e resolução de problemas de governança e avaliação dos ativos de dados existentes (DAMA, 2017).

Um dos desafios da governança e gestão de dados é em sua aplicação prática, onde se observam problemas pela complexidade existente nas organizações. Esses desafios geralmente giram em torno da percepção de valor nos diferentes níveis da organização, responsabilização e formação de autoridade competente para com os dados, a formação de silos de dados, as mudanças tecnológicas constantes e ambientação política que dificulta transformações (BENFELDT; PERSSON; MADSEN, 2019).

Dentro do aspecto de responsabilização existe uma dificuldade inerente a esse processo, que ainda não se possui comum entendimento claro na literatura, que é o papel do responsável pelos dados, conhecido como *data owner*. Suas atribuições e como se determina esse papel dentro da organização é atualmente uma difícil tarefa e que necessita de aprofundamento (VILMINKO-HEIKKINEN; PEKKOLA, 2019).

Essa dificuldade, assim como outras, podem criar problemas de implementação de modelos, principalmente se tratando de um conceito abstrato de propriedade, responsabilização e escopo no processo de governança de dados, e é um pilar importante para que exista um melhor planejamento de governança nas organizações (ABRAHAM; SCHNEIDER; BROCKE, 2019).

Portanto mediante este cenário exposto, essa pesquisa busca explanar o conceito do *data owner*, suas atribuições, responsabilização e escopo dentro do contexto de aplicação da governança de dados em organizações, tendo como questão norteadora: **Como organizações podem determinar a alocação papel de autoridade do *data owner* e quais as suas responsabilidades?**

1.1 OBJETIVOS

Nessa seção serão abordados os objetivos que guiam essa pesquisa, subdividido em objetivo geral e objetivo específico apresentando a finalidade em contexto desse trabalho e suas intenções específicas que a serem atingidas.

1.1.1 Objetivo geral

O objetivo geral dessa pesquisa é identificar como organizações podem determinar a definição do papel de autoridade do *data owner* e quais as suas responsabilidades.

1.1.2 *Objetivos específicos*

Os objetivos específicos dessa pesquisa são dentro de suas fases metodológicas incorporar a aplicabilidade de estudos significativos sobre a governança de dados que foram realizados na prática de gestão de dados e entender a aplicabilidade desses estudos para com o papel do *data owner* buscando observar os seguintes aspectos:

- Descrever a evolução diferenciação e importância da gestão de dados na figura da governança, autoridade e controle.
- Identificar o exercício da governança de dados dentro das organizações e sua problemática com relação aos papéis e responsabilidades adotados.
- Reunir os diversos ângulos apresentados na literatura sobre o papel do *data owner* no processo de formulação e práticas de governança de dados.
- Esclarecer as melhores práticas existentes de determinação do papel do *data owner* dentro de organizações, assim como suas responsabilidades.

1.2 JUSTIFICATIVA

Para a apresentação da justificativa e se obter um entendimento compreensivo das intenções e motivações desse trabalho foi realizada uma subdivisão nos aspectos científico e socioeconômico.

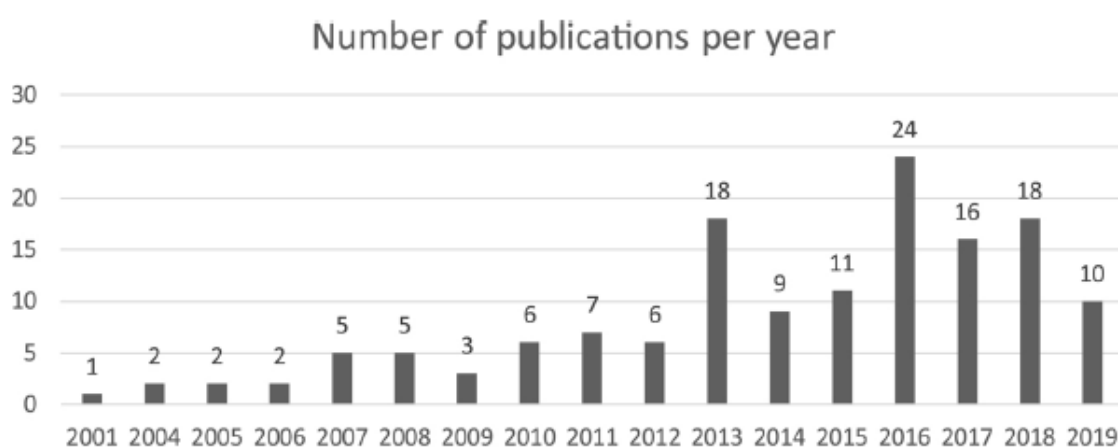
1.2.1 *Aspecto Científico*

A gestão e governança de dados aparece dentro do ambiente científico nas mais diversas áreas que a veem como importante necessidade de aprofundamento da sua conceituação, problematização e modelagem. Porém, a literatura com relação a governança de dados apresenta um intrincado desenvolvimento conceitual e de aplicações práticas, onde o atual estado da arte da literatura ainda promove

questionamentos e problemas essenciais a serem resolvidos. Esses problemas se expandem em diversas áreas de pesquisa, como por exemplo os mecanismos de governança, o escopo da governança dentro dos mecanismos organizacionais, a contextualização dos antecedentes à governança de dados, e as consequências ou efetividade da governança aplicada (ABRAHAM; SCHNEIDER; BROCKE, 2019).

As publicações relevantes especificamente na área de governança de dados expandiram em número durante os últimos anos, principalmente devido ao aumento dos estudos de caso, em análises métricas da literatura já se observa essa tendência.

FIGURA 1 - PUBLICAÇÕES RELEVANTES EM GOVERNANÇA DE DADOS



FONTE: Abraham, Schneider e Brocke (2019, p.427)

A governança de dados ganhou principalmente maior atenção e relevância nas áreas de *Big Data* e *Big, Open and Linked Data* (BOLD) e consequentemente na implementação de sistemas de *machine learning* e outras formas de inteligência artificial em larga escala, onde nessa nova perspectiva existe a inerente necessidade de estruturação, interoperabilidade e qualidade de dados em sua vasta quantidade e diversidade (JANSSEN et al., 2019).

Existe também, por outro ponto de vista, o aspecto dos benefícios que esses modelos e técnicas trazem para a própria ciência como um todo. O movimento atual de acesso aberto, com o compartilhamento, a reutilização e a redistribuição de dados científicos conhecido pelos conceitos de Ciência Aberta e Dados Abertos, que se consolida em muitos países pelo mundo, e a prioridade dentro dessa movimentação

tem se focado no processo de abertura dos dados científicos (SANTOS; ALMEIDA; HENNING, 2018).

Esse processo de abertura envolve profundamente a gestão de dados, principalmente se tratando das infraestruturas tecnológicas que suportam esses dados científicos, e a partir delas o processo de gestão e governança que permeia e que é um dos principais condutores da descoberta de conhecimento e da inovação dentro do ciclo de vida dos dados, desde sua coleta, armazenamento, compartilhamento e reutilização (WILKINSON et al., 2016).

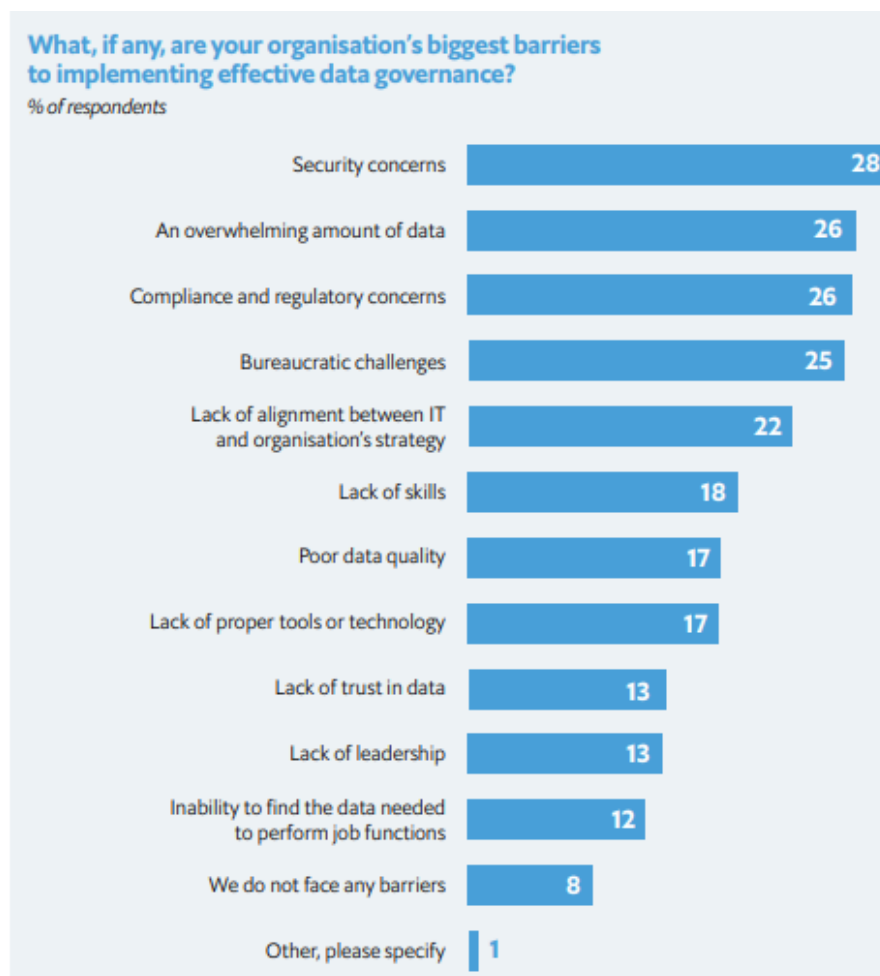
1.2.2 Aspecto Socioeconômico

O desenvolvimento social há muito é entendido como tendo relação direta com o desenvolvimento científico e tecnológico, onde seu conhecimento por natureza possui grande autoridade e permite a liderança de políticas de mudanças sociais necessárias (YEARLEY, 2015).

A revolução tecnológica atual traz consigo um desafio sem precedentes, onde bilhões de pessoas, no mundo inteiro produzem dados estruturados e não-estruturados a todo momento, em enorme e crescente quantidade, por meio do uso de aparelhos digitais e internet (KHAN et al., 2014). Na indústria, o desenvolvimento de tecnologias de integração de sistemas, computação em nuvem, internet das coisas, *big data & analytics* entre outros, que concede aos dados uma relevância que se expande cada vez mais dentro dessas organizações (SESI, 2020).

Em pesquisa realizada pelo grupo The Economist Intelligence Unit (2018) onde foram entrevistados mais de 500 executivos de empresas dos mais diversos ramos do mercado, e demonstra que existe um entendimento comum da importância da governança de dados dentro das organizações, sendo relacionada como um conceito de fundação ou base das organizações. Assim como é de entendimento comum que uma governança, quando realizada com rigor e qualidade auxilia as organizações a criarem vantagens competitivas por meio dos seus dados revelando pontos de melhoria como mercados inexplorados e despesas desnecessárias.

FIGURA 2 - BARREIRAS PARA IMPLEMENTAÇÃO DA GOVERNANÇA



FONTE: The Economist (2018, p.11)

Um aspecto interessante dessa pesquisa é a visão desses empresários e gestores com relação as maiores barreiras para sua implementação efetiva dentro das organizações. Observa-se os desafios mais votados, de segurança e quantidade de dados que são óbvios, mas denotam o atraso na implementação da governança, e as demais barreiras que em grande parte são relacionadas a figura do *Data Owner*, com relação a complexa tarefa de autoridade e controle dos dados, do processo e dos desafios burocráticos, do alinhamento com as diversas áreas incluindo as de tecnologia e da liderança nesse processo.

Segundo Manyika et al. (2011), o uso efetivo e massivo dos dados tem o potencial trazer uma nova onda de crescimento produtivo e de transformar economias. Tendo o poder de ganhos operacionais, direcionamento estratégico, criação de novos produtos e descobrimento de novos mercados. Portanto deve ser a base de

estratégias competitivas de organizações, onde novos competidores devem focar em atrair colaboradores com habilidades e conhecimentos nas diversas áreas de dados.

O aspecto socioeconômico é um dos pilares dessa pesquisa por ser fonte mais clara do problema e das possíveis soluções para o pesquisador. Na prática econômica e dentro das organizações é o ambiente onde muitas estratégias se aplicam, onde se toma riscos com novas ideias, e adotam-se modelos que entram em uma forma de seleção natural dentro de cada organização, e onde os resultados comprovam ou contestam os modelos propostos de acordo com as necessidades e demandas de cada organização.

Existindo um intrincado desenvolvimento conceitual e de aplicação no aspecto científico, no âmbito organizacional isso é potencializado, sendo uma dificuldade maior de conceituação dos termos e aplicação de qualquer modelo, principalmente levando em consideração a extrema complexidade das organizações, e expondo que o estudo da governança geralmente se dá nos aspectos isolados e específicos, quando na prática a simples indefinição da responsabilização de forma compreensiva já encaminha os projetos de implementação de governança para um ambiente desfavorável (BENFELDT; PERSSON; MADSEN, 2019).

2 REFERENCIAL TEÓRICO

Nessa seção é apresentado o referencial teórico para fundamentar a pesquisa, para o desenvolvimento desse trabalho foram levantados conceitos teóricos relativos à gestão de dados e governança de dados.

2.1 GESTÃO DE DADOS

Dados são observações, coleções de evidências, representações de fatos ou estado do mundo por pessoas ou tecnologias, e se diferenciam da informação pela sua abstração de contexto, sua relevância, seu propósito e, portanto, pela intervenção humana. Os dados também não simplesmente existem, e sim são criados a partir de pessoas ou tecnologias, e possuindo um suporte, estrutura, armazenamento e, portanto, necessitam também de uma gestão apropriada (DAVENPORT, 1998; SORDI, 2015; DAMA, (2017)).

A gestão de dados ou *data management* (DM) é definida por DAMA (2017) como o desenvolvimento, a execução e supervisão de planos, políticas, programas e práticas que entregam, controlam, protegem e aprimoram o valor dos dados como ativos durante seu ciclo de vida. Sua aplicação prática nas organizações atravessa tecnologias, processos e pessoas, e também depende de diversas áreas de conhecimento como:

- Governança de Dados
- Integração e Interoperabilidade
- Modelagem e Design de Dados
- Segurança de Dados
- Armazenamento e Operações
- Metadados
- Qualidade de Dados
- Arquitetura de Dados
- Referência e Dados Mestres

Esse conceito se diferencia da governança de dados onde a governança de dados se refere a formulação de políticas, práticas e decisões que devem ser tomadas, assim como quem deve tomar essas decisões, enquanto a gestão de dados

se preocupa com a execução de tais políticas, práticas e decisões (ABRAHAM; SCHNEIDER; BROCKE, 2019).

2.1.1 Dados como Ativo Organizacional

Organizações cada vez mais consideram seus dados como ativos, recursos econômicos, e a gestão de dados tem um importante papel no processo de implementação e transformação dos dados como ativos organizacionais (KHATRI; BROWN, 2010; PERRONS; JENSEN, 2015).

Esse processo vem de forma similar e talvez por extensão (PIERCE; DISMUTE; YONKE, 2008) a um processo já conhecido de transformação da informação como ativo estratégico intangível, para os diversos tipos de organizações, onde as informações passaram a necessitar de autoridade, controle, responsabilização, políticas, obrigações, avaliações de risco, segurança etc., e portanto emerge-se disso a importância de gerenciá-los de forma estratégica (MCGEE; PRUSAK, 1994; STEVENS; CARALLI; WILLKE, 2005; EVANS; PRICE, 2012; VISCUSI; BATINI, 2014).

Da mesma forma que a informação, os dados devem cada vez mais ser reconhecidos como ativos nas organizações e receberem também o mesmo tratamento, investimento e atenção advindos dessa classificação (FISHER, 2008). Como ativos organizacionais, os dados são tratados de forma similar aos outros ativos organizacionais e necessitando de gestão de ativos para garantir a manutenção de suas propriedades de geração de valor (REDMAN, 2008).

Portanto, dados passam a ter as necessidades gestão como criação de políticas, autoridade, controle, custódia, e integração no processo de gestão de ativos, para aplicações de melhores práticas tem como objetivo a maximização do valor desses dados dentro das organizações (MCGILVRAY, 2006; CARRETERO et al., 2017).

Dentro desse aspecto, a gestão de dados inclui desde as habilidades de gerar decisões consistentes sobre a estratégia de extração de valor dos dados, a aplicação técnica e prática dessas estratégias e a supervisão desse processo, portanto atravessa as diversas áreas de conhecimento técnico e de negócios e as responsabilidades dentro dessas áreas são compartilhadas (DAMA, 2017).

2.1.2 Ciclo de Vida dos Dados

O ciclo de vida dos dados ou *data lifecycle* entende-se como processos de criação ou obtenção de dados, da movimentação, da transformação, do armazenamento, da facilitação, do compartilhamento, da aplicação e do seu descarte. (MCGILVRAY, 2006). Durante esses processos os dados podem também passar por diversos processamentos, limpezas, melhorias, transformações, combinações etc.

Portanto, o ciclo de vida é o mapeamento de onde a gestão de dados deve ocorrer, e também para identificar problemas como a distribuição dos dados em diversos sistemas, por diversas áreas e autoridades, intraorganizacionais ou interorganizacionais, e que, portanto necessitam de um olhar minucioso para entendimento dos dados por diversas entidades.

Segundo Carretero et al. (2017) a governança de dados com foco no ciclo de vida dos dados se relaciona com a identificação dos processos de negócio que utilizam os dados. Portanto, o ciclo de vida deve ser relacionado com cada ciclo de vida do produto de negócio respectivo, pois cada produto terá sua necessidade ou uso de dados de forma diferenciada.

Diferentes tipos de dados têm diferentes necessidades de gestão e, portanto, diferentes tipos de ciclo de vida. Em se tratando de *Big Data*, os dados possuem características especiais como seu volume e escala, onde existe a geração e coleção de dados de forma exponencial. A sua variedade, onde podem existir os diversos tipos de dados (estruturados e não-estruturados). Sua velocidade, onde a coleta desses dados deve ocorrer de forma rápida, portanto existindo a preocupação temporal de coleta e tratamento. Outras características importantes também se tratando do seu valor econômico em quantidade, que necessita da rápida captura, descobrimento e análise. (LANEY, 2001; GANTZ et al. 2011).

Portanto, pela sua proporção, que vai muito além da capacidade de bancos de dados tradicionais de coleta, armazenamento, gestão e análise, *Big Data* possui uma necessidade especial de mapeamento e controle de seu ciclo de vida, é um problema em essência pela complexidade dos sistemas e processos envolvidos, somado também a complexidade desses dados em larga escala, variedade e velocidade. (DAKI et al., 2017).

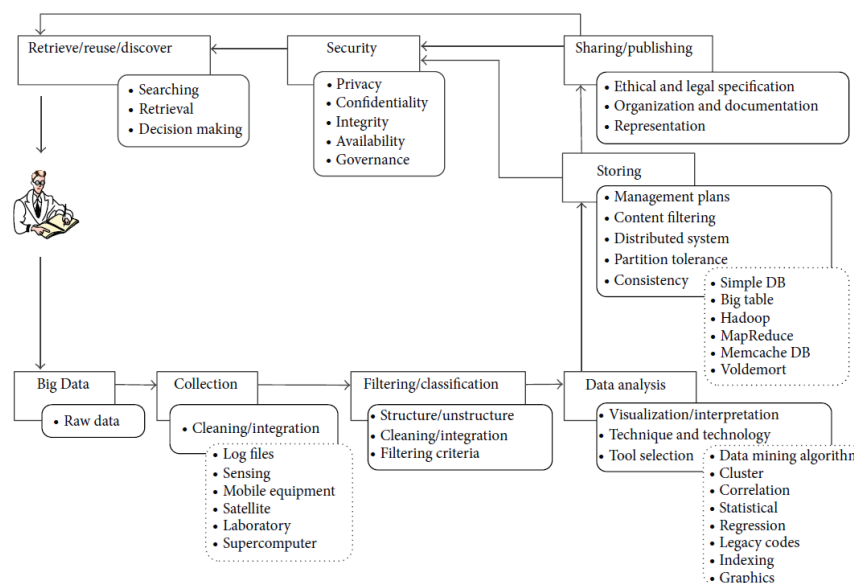
O processo de análise desses dados ou *Big Data Analytics* (BDA), é definido como a investigação desses dados para descobrir correlações, padrões escondidos

pela forma de distribuição dos dados, que podem gerar informações de preferências de consumidores, das tendências de mercado, problemas de produção industrial, e auxiliar as organizações e a sociedade na sua tomada de decisão. (ABELL et al., 2017).

Para tanto, é necessário o mapeamento do ciclo de vida de BDA, que pode estar relacionado ao próprio ciclo de vida dos produtos analisados (REN et al., 2019), desde sua criação e a arquitetura desses dados, sua interconexão relevante com outros dados existentes no mesmo processo de produção, as tecnologias e sistemas envolvidos, os cuidados com segurança desses dados, quais áreas e fontes envolvidas em todo esse processo compreendendo de ponta a ponta a corrente de análise desses dados.

Segundo Khan et al. (2014), o volume desses dados e a sua velocidade de criação causa que proporcionalmente, esses dados individualmente têm um decréscimo de valor conforme algumas características como idade, tipo, qualidade e riqueza, propondo um modelo geral com as tecnologias e terminologias utilizadas na área de BDA.

O modelo proposto (ver Figura 3) mapeia o processo desde a criação dos dados brutos por meio de mensagens, *logs*, sensores, equipamentos móveis, satélites, laboratórios, e até mesmo mensagens pela internet. Assim como sua coleta, filtragem e classificação, que integra, que enriquece, que limpa e estrutura esses dados para uma análise, e a partir dessa análise esses dados passam pela fase de armazenamento, o compartilhamento ou publicação, a segurança e governança e recuperação para reuso desses dados.

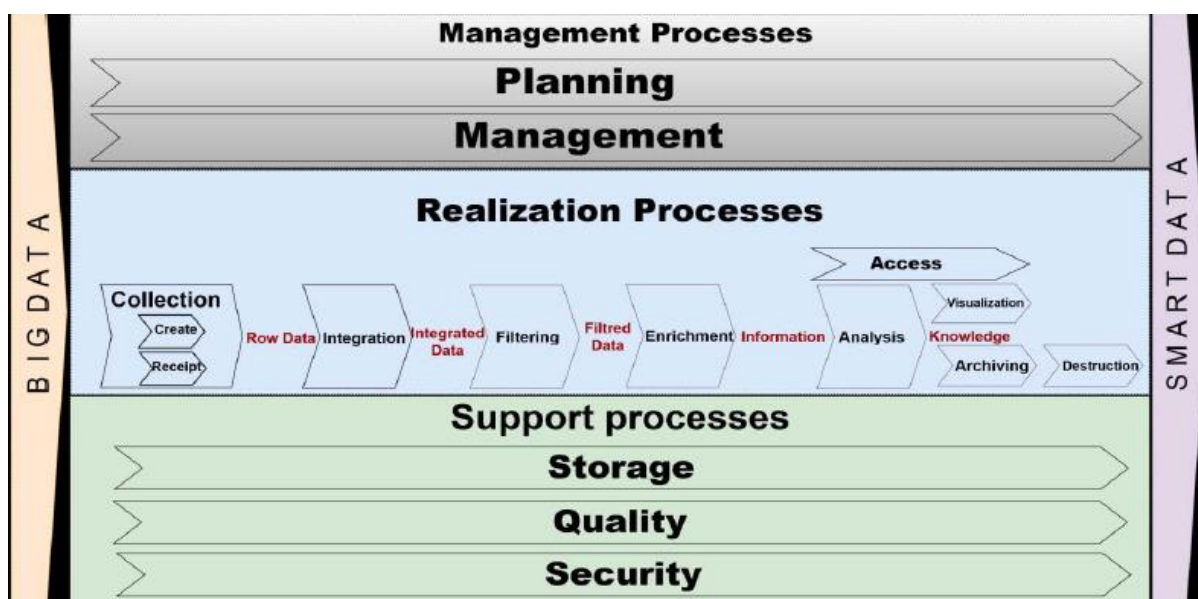
FIGURA 3 - PROPOSTA DE CICLO DE VIDA DE *BIG DATA ANALYTICS*

FONTE: KHAN et al. (2014, p.9)

Uma porção significativa dos estágios desse modelo geral de ciclo de vida de *Big Data* têm como componente essencial a gestão de dados. Organizações enfrentam problemas com relação a criação, a gestão e manipulação do fluxo crescente e rápido de dados, a modelagem desse ciclo de vida permite sua escalabilidade com acurácia devido as complexidades de sistemas, tecnologias e processos envolvidos (KHAN et al., 2014).

Segundo Arass e Souissi (2018), em modelo proposto relativo ao ciclo de vida de *Big Data Analytics* para criação do que chama da *Smart Data* (Figura 4) utilizando-se da teoria de gestão informacional possui também uma reflexão importante e presente em outros modelos similares, que é o conceito do descarte de dados, seu modelo passa por planejamento preliminar, a gestão operacional, a coleta dos dados, a integração das fontes, a filtragem dos dados, o enriquecimento da qualidade dos dados, a análise e acesso, e por fim o arquivamento ou descarte.

FIGURA 4 - CICLO DE VIDA PROPOSTO SMART DATA LIFECYCLE



FONTE: Arass e Souissi (2018, p.84)

Portanto, para a gestão de dados de forma efetiva existe a necessidade de planejamento e estratégia com relação ao ciclo de vida dos dados existentes nas organizações, e esse ciclo de vida pode ser gerenciado estrategicamente com a visão de como a organização irá utilizar os seus dados da melhor forma.

O ciclo de vida dos dados envolve também a linhagem de dados, que são os caminhos que os dados se movimentam dentro de sistemas e processos, que partem de um ponto de origem para um ponto de uso. Linhagem de dados é um processo de documentação das origens, das movimentações, das transformações, pontos de acesso e de uso dos dados. O processo de mapeamento de linhagem de dados é essencial para o mapeamento do seu ciclo de vida, principalmente em seus pontos críticos de criação e uso (DAMA, 2017).

É comum que organizações produzam diversos tipos e grandes quantidades de dados que nunca são utilizados, a gestão do ciclo de vida dos dados não tem como preocupação o mapeamento de todo e qualquer processo de criação e utilização de dados, mas sim em uma gestão focada nos dados críticos das organizações, reduzindo a redundância, obsolescência e trivialidade (AIKEN, 2014).

2.2 GOVERNANÇA DE DADOS

Segundo DAMA (2017), a governança de dados ou *data governance* (DG) é definida como o exercício de autoridade e controle sobre os dados, sua prática envolve planejamento, a execução e monitoramento de estratégias, de políticas, de modelos, de conformidade e problemas relacionados a gestão dos dados.

Mesmo que involuntariamente, as decisões são tomadas pelas organizações sobre seus dados, portanto, a governança é o controle formal sobre essa tomada de decisão (SEINER, 2014). Para tanto, é necessário a criação de modelos para formalizar o tomador de decisão, sua responsabilização e seu escopo de decisão, assim como prover regras para o uso e tratamento dos dados das organizações, e permitindo a formalização de procedimentos, políticas, padrões e monitoramento de conformidade (WEBER; OTTO; ÖSTERLE, 2009; KHATRI; BROWN, 2010; NASCIO, 2008).

No ponto de vista de Dyché e Levy (2008) a governança de dados é um processo fundamental para a gestão efetiva desses dados, e deve se espalhar pela organização, portanto atravessando níveis organizacionais e linhas de negócio. Assim como sua aplicação nas organizações geralmente está atrelada a uma complexidade de sistemas, dos processos e pessoas, onde a busca pela criação de modelos e aplicações de forma sistemática, otimizada e reproduzível em estudos de caso ainda é um desafio (ABRAHAM; SCHNEIDER; BROCKE, 2019).

Governança implica em aplicação de poder direcionada a processos e interações das organizações com regulamentos, normas, leis, direcionamentos e costumes dentro das organizações (BEVIR, 2012). Portanto, seu alinhamento com a estratégia organizacional é uma necessidade. A governança não pode ter como objetivo simplesmente existir, ou aparentar simplesmente existir, a clareza do seu relacionamento, com a finalidade de resolução de problemas dentro das organizações, é essencial para a sua adoção, mudança de comportamento e cultura, se mantendo sustentável no longo prazo e incorporada aos processos organizacionais (DAMA, 2017).

Segundo o *Data Governance Institute*, a fundação de uma governança de dados gira em torno de liderança, estratégia e parceria de áreas de tecnologia que são guiadas pelas necessidades das áreas de negócio das organizações, que devem possuir responsabilidades claras e compartilhadas nos seus diferentes níveis. Sua coordenação e operações devem ser baseadas em modelos e normas que definem

as interações e responsabilidades de cada colaborador, e criados princípios básicos a serem seguidos como essência e referência, tendo como princípios a integridade, transparência, auditoria, controle, equilíbrio, padronização e gestão de mudanças.

2.2.1 Modelos e Estrutura de Governança

Segundo Abraham, Schneider e Brocke (2019), para o estudo dos modelos de DG é importante levar em consideração o escopo dos dados e suas necessidades específicas para cada caso, e separando o tratamento de dados tradicionais e *Big Data*, onde todo projeto de implementação de governança de dados deve levar em conta essa diferença e seu foco.

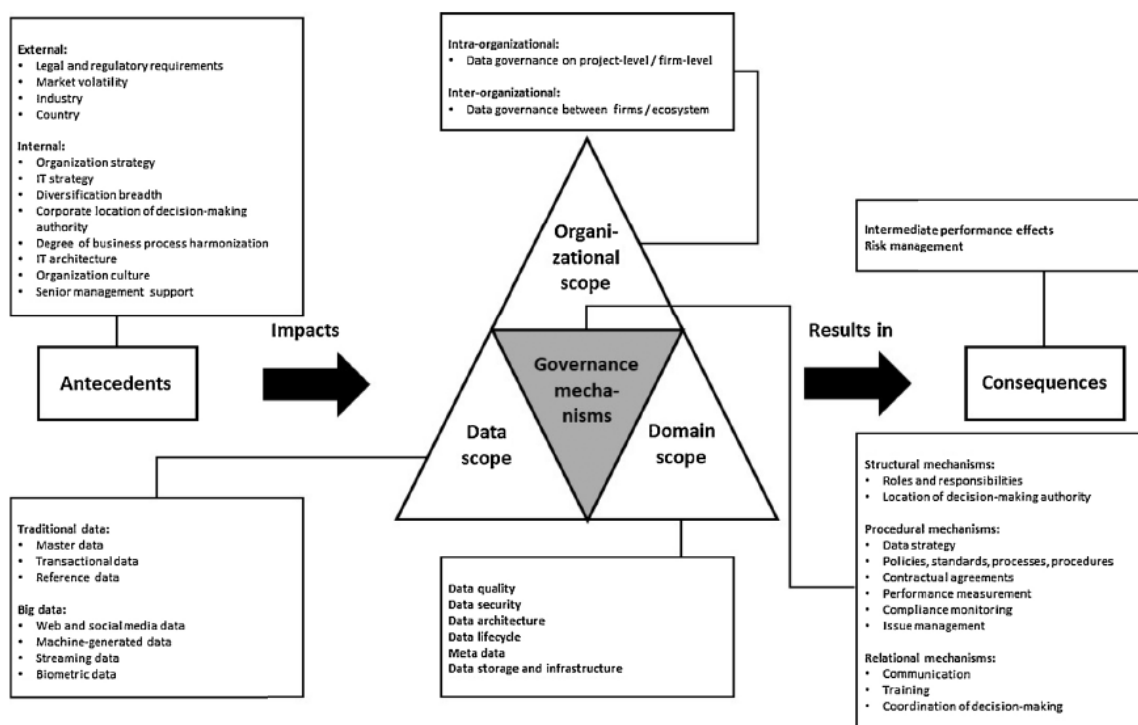
Dados tradicionais são dados estruturados geralmente separados em dados mestres, transacionais e referenciais. Alguns exemplos de dados mestres são mapeamentos de clientes, funcionários, finanças, produtos, materiais, locais, dentre outros. Dados transacionais são relativos a transações financeiras, transporte, logística, dentre outros. Já os dados referenciais são relativos a dados que devem ser unificados internamente na organização como códigos de produtos, numeração de equipamentos, nomes de processos, siglas e abreviaturas etc. (DREIBELBIS et al., 2008).

Em se tratando de *Big Data* existe um nível maior de complexidade principalmente por sua variedade, que, portanto, permite os mais diversos tipos de dados. Dados estruturados similares aos tradicionais, porém em grande volume e velocidade. Dados semiestruturados como *logs* de sistemas, dados de sensores, dentre outros. E dados não-estruturados como áudios, imagens, vídeos (LANEY, 2001). Alguns exemplos de *Big Data* são estruturas de dado gerados por pessoas nas redes sociais e internet em geral, dados gerados por máquinas nas organizações e na indústria, dados financeiros gerados por instituições bancárias, dentre outros. (TALLON, 2013; REN et al., 2019).

Existe também os escopos de domínio e organizacionais desses dados, onde o domínio é relativo as áreas que tratam da gestão desses dados como já citadas anteriormente, por exemplo as áreas de qualidade de dados, segurança de dados, arquitetura e armazenamento de dados etc. E o escopo organizacional é relativo à unidade controladora dos dados, são divididos em intraorganizacionais, ou seja, de dentro de uma única organização e governados a nível de projeto ou a nível corporativo, ou interorganizacionais que são dados compartilhados entre duas ou mais

organizações, que foram colaborações, parcerias ou ecossistemas de dados. (TIWANA et al., 2014; DAMA, 2017). Abraham, Schneider e Brocke (2019) em seu modelo conceitual apresentado na Figura 5, buscam representar esses conceitos relativos aos mecanismos de DG com seus domínios e escopos.

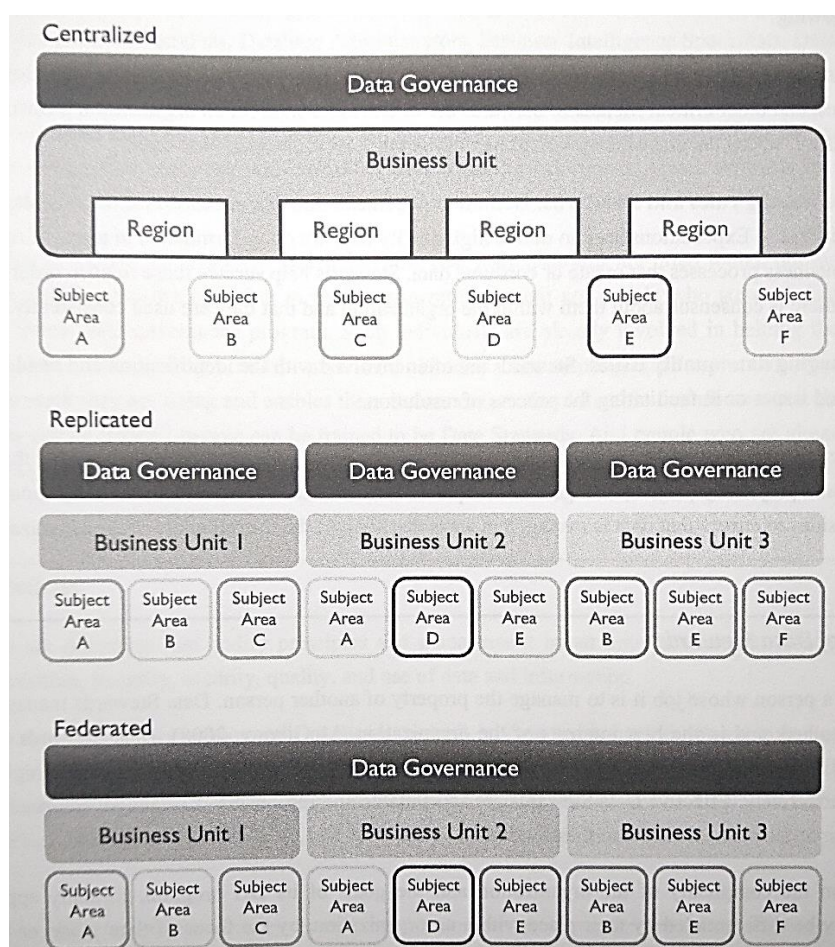
FIGURA 5 - CONCEITOS RELATIVOS A MODELOS DE GOVERNANÇA



FONTE: Abraham, Schneider e Brocke (2019, p.429)

Tendo em perspectiva a complexidade de escopo dos dados, existe ainda diferentes tipos de estrutura onde elas podem se aplicar. Segundo Ladley (2012), se tratando de estruturas de governança existem três tipos, estruturas centralizadas, replicadas e federadas. Dentro de uma estrutura centralizada, uma área de governança de dados controla e gerencia todas as atividades em uma estrutura única e central. A estrutura replicada reparte a autoridade de uma mesma estratégia de governança nas diversas áreas de negócio que tomam suas decisões de forma independente. Em uma estrutura federada possuindo níveis adicionais de autoridade, uma área centraliza a estratégia de DG e empodera as diversas unidades de negócio da organização a trabalharem com seus diferentes tipos, complexidades e escopo de dados (DAMA, 2017).

FIGURA 6 - ESTRUTURAS DE GOVERNANÇA DE DADOS



FONTE: DAMA (2017, p.75)

2.2.2 Data Stewards e Data Owner

Segundo McGilvray (2006), o conceito de *data steward* é relativo à gerência de dados como ativos em nome de outros, e seu papel representa os interesses estratégicos da organização. O papel do *data steward* é comparável ao de um agricultor que toma conta de uma terra gerenciando a plantação e colheita, sem necessariamente ser o dono da terra, assim como o *data steward* não é dono dos dados. (BITTERER; NEWMAN, 2007).

Segundo Cheong e Chang (2007), *data steward* tem um papel de liderança de negócios, e deve possuir conhecimentos específicos dos requerimentos de negócios para com seus dados, e que devem traduzir esses conhecimentos para as áreas técnicas que trabalham em paralelo. Esse conceito também pode ser subdividido em um trabalho relacionado ao negócio, *business data steward*, ou

relacionado a área de conhecimentos de tecnologia da informação, *technical data steward*, as duas são responsabilidades que trabalham em paralelo e gerenciando os mesmos dados.

Para Weber, Otto e Österle (2009), *technical data stewards* são responsáveis pela padronização de elementos, definições, formatos, e detalhamentos de perspectiva técnica para os fluxos de dados. Já o *business data steward* é responsável por governar os padrões e políticas de dados em seu domínio dentro da perspectiva da sua área de negócios.

O *Data Owner* está, para DAMA (2017), dentro da figura de um *business data steward*, tendo o papel de autoridade e decisões sobre os dados em seu domínio. Necessitando, portanto, ter conhecimento específico do negócio relativo aos dados que governa, tendo decisão final sobre seu domínio de dados, sua manutenção e desenvolvimento (HEIKKINEN; PEKKOLA, 2019).

A identificação do *data owner* é um passo essencial principalmente para a gestão de dados, a execução das políticas, e deve ser realizada reconhecendo os processos e necessidades de negócio, e nisso o *data owner* é peça chave. A indefinição desse papel gera desalinhamentos e conflitos que podem dificultar a implementação de governança dos dados e aplicação da gestão de dados dentro das organizações (BENFELDT; PERSSON; MADSEN, 2019, HEIKKINEN; PEKKOLA, 2019).

3 METODOLOGIA

Nessa seção será apresentado o encaminhamento metodológico da pesquisa, que é subdividido em caracterização da pesquisa e etapas do procedimento metodológico da pesquisa.

3.1 CARACTERIZAÇÃO DA PESQUISA

Considerando que esse trabalho tem como objetivo prático a geração de conhecimentos para aplicação e solução de problemas específicos dentro da área de governança de dados, esse trabalho tem como sua natureza aplicada.

O problema de pesquisa escolhido para esse trabalho possui um aspecto estritamente conceitual, subjetivo e dinâmico, a análise do papel do *data owner* é descrito de diversas formas na literatura, e o seu aprofundamento terá como base uma abordagem exclusivamente qualitativa.

Segundo Gil (2002), a pesquisa é um procedimento que tem como objetivo proporcionar respostas aos problemas propostos de forma racional e sistemática por meio de procedimentos e métodos científicos. Para Silva e Menezes (2005), o conceito de pesquisa de forma simples, seria a procura de resposta para indagações propostas. Sendo, portanto, um conjunto de propostas e ações que tem como fim encontrar soluções para um problema baseando-se em procedimentos racionais e sistemáticos.

A pesquisa também deve ser caracterizada nos diversos pontos de vista, compreendendo sua natureza, forma de abordagem, objetivos e procedimentos técnicos (SILVA; MENEZES, 2005). A Figura 7 tem como objetivo refletir essa caracterização para propósito desse trabalho.

FIGURA 7 - CARACTERIZAÇÃO DA PESQUISA



FONTE: O autor (2023)

Para fins de objetivos, se tratando de um conceito existente na literatura, e um fenômeno de problemáticas relativas a esse conceito, esse trabalho tem como objetivo a explicitação do problema e construção de hipótese a partir das teorias analisadas, portanto será uma pesquisa descritiva e exploratória.

Do ponto de vista da abordagem técnica, pensando em como melhor resolver o impasse da conceituação e aplicação prática do problema escolhido, especialmente devido à dificuldade de estudo da implementação da governança por ser uma atividade estratégica das organizações, foi escolhida, portanto, a pesquisa bibliográfica com revisão integrativa.

3.2 MÉTODO E PROCEDIMENTOS

Segundo Mendes (2008), a revisão integrativa é um modelo que inclui a realização de análise de pesquisas relevantes que auxiliam a tomada de decisão, e que possibilita a síntese do estado da arte de um tema ou conhecimento específico, apontando possíveis lacunas desse conhecimento que necessitam de preenchimento a ser realizado por meio de novos estudos, esse método portanto permite sintetizar conclusões gerais sobre uma área específica de estudo.

Para Ercole (2014), a variedade possível na composição amostral de uma revisão integrativa assim como sua múltipla finalidade permite resultados que

visualizam um quadro completo de conceitos complexos, teorias ou problemas existentes

Segundo Souza (2008), a revisão integrativa surge como um processo metodológico que proporciona síntese do conhecimento juntamente com a aplicabilidade de resultado de estudos práticos, sendo a mais ampla abordagem metodológica e permitindo a compreensão completa de um fenômeno analisado, e combinando dados da literatura teórica e empírica.

Para propósito desse trabalho serão seguidos os procedimentos descritos por Souza (2008) para realização de uma revisão integrativa, que incluem fases de elaboração de pergunta norteadora, a busca ou amostragem na literatura, a coleta de dados, a análise crítica dos estudos incluídos, a discussão dos resultados, para enfim, a apresentação final de uma revisão integrativa.

Com propósito de identificar conceitos dentro da literatura de governança de dados, se utilizando de processo metodológico de revisão integrativa da literatura descrita por Souza (2008), tema e problema definido como o processo de determinação, a identificação e alocação da figura de autoridade do *data owner* para a governança de dados e suas devidas responsabilidades.

3.3 BUSCA NA LITERATURA

Para a realização de busca na literatura, as bases de dados escolhidas foram Scopus, *Web of Science* (WOS) e IEEE Xplore. Sendo buscados especificamente artigos, e como estratégia de busca com palavras-chave, serão utilizados os descritores "*data governance*", "*data management*" ou "*data mesh*", e também necessariamente o descritor "*data owner*", para obtenção de artigos relevantes, assim como visto a atualidade da discussão, durante os períodos de 2020 a 2023.

O processo de busca foi realizado observando que o critério de elegibilidade dos resultados é a partir de uma análise do título, resumo e palavras-chave do artigo se contém não somente os termos descritores relacionados, mas também sua relevância dentro da perspectiva de análise. Resultados de artigos avaliados e escolhidos como relevantes são aqueles que após uma leitura técnica se observou um contexto de importância por estudarem casos específicos sobre a questão da responsabilização e controle dos dados no processo de gestão de governança.

QUADRO 1 – BUSCA E COLETA DE ARTIGOS NA LITERATURA

Base	ScienceDirect	Web of Science	IEEE Xplore
Website	https://www.sciencedirect.com/	www.webofscience.com	ieeexplore.ieee.org
Busca Avançada	Título, Resumo, Palavras-chave	Título, Resumo, Palavras-chave	Título, Resumo, Palavras-chave
Descritores na busca	("data management" OR "data governance" OR "data mesh") AND ("data owner")	("data management" OR "data governance" OR "data mesh") AND ("data owner")	("data management" OR "data governance" OR "data mesh") AND ("data owner")
Tipo de publicação	Artigo	Artigo	Artigo
Período	2020-2023	2020-2023	2020-2023
Data da busca	05/01/2023	16/02/2023	07/01/2023
Resultados	51	19	5
Artigos Seleccionados	6	4	2

FONTE: O autor (2023).

Após a coleta dos artigos foi realizada uma seleção dos documentos com leitura dos resumos, e observando que o contexto desses artigos relacionado ao tema como supracitado, assim como sobreposição ou coincidência da existência do mesmo artigo em múltiplas bases, sendo esses contabilizados nos resultados de casa base e selecionados apenas de uma. Foi também realizada uma separação dos artigos devido aos seus temas em comum, problemas e tendencias encontradas na literatura.

3.4 COLETA DE INFORMAÇÕES DE CONFERÊNCIA E OUTROS

Em um processo de coleta de informações relevantes e realização de uma proposta de análise compreensiva sobre a área de gestão e governança de dados é necessária uma busca complementar de fontes de informações confiáveis e relevantes sobre o tema.

Segundo Booth, Colomb e Williams (2008), a participação em conferências da área de estudo apresenta uma oportunidade para verificar e aprender novos desenvolvimentos de ideias que estão sendo realizadas no presente e futuro, e aprender sobre inovações e tendências, para conectar pesquisadores e

colaboradores, assim como apresentar e receber feedback de ideias e pesquisas sendo realizadas.

Para esse trabalho, uma dessas fontes utilizadas foi a coleta de informações na participação de conferência da área de gestão e governança de dados *Enterprise Data World Digital* realizada no ano de 2022, visto que a área da gestão e governança de dados especialmente no campo de Big Data Analytics é ainda muito recente e as discussões sobre melhores abordagens, casos de uso e tendência aparecem em uma diversidade de documentos e livros. Algumas dessas recentes publicações sobre o tema que são citados por múltiplas fontes também foram utilizados dentro do escopo de análise.

Foram coletados, portanto, informações de conferências e documentos que serão posteriormente analisados conjuntamente com a busca na literatura acadêmica objetivando uma ampla observação dos modelos existentes na literatura e casos apresentados, e combinando a teoria com a realidade prática do mercado e movimentações atuais de possíveis modelos e soluções para a governança de dados no aspecto do *data owner*.

4 RESULTADOS

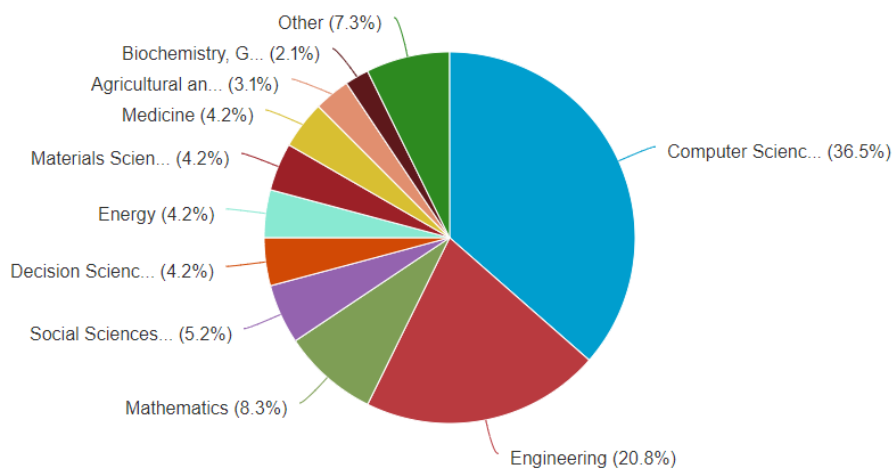
Nessa seção serão apresentados os resultados da pesquisa seguindo o procedimento apresentado anteriormente, após a realização da busca na literatura e coleta de informações pertinentes, foi realizado o procedimento de análise crítica e discussão dos resultados levantados.

4.1 ANÁLISE DOS RESULTADOS: EVOLUÇÃO DA LITERATURA

Importante ressaltar que a evolução e importância da gestão de dados na figura da governança em organizações vem em uma crescente discussão em andamento e muitos conceitos abordados são também recém-criados, problemática essa que pode possuir diversos ângulos de abordagem, essa análise tem como foco explorar a figura da governança nesses diferentes aspectos.

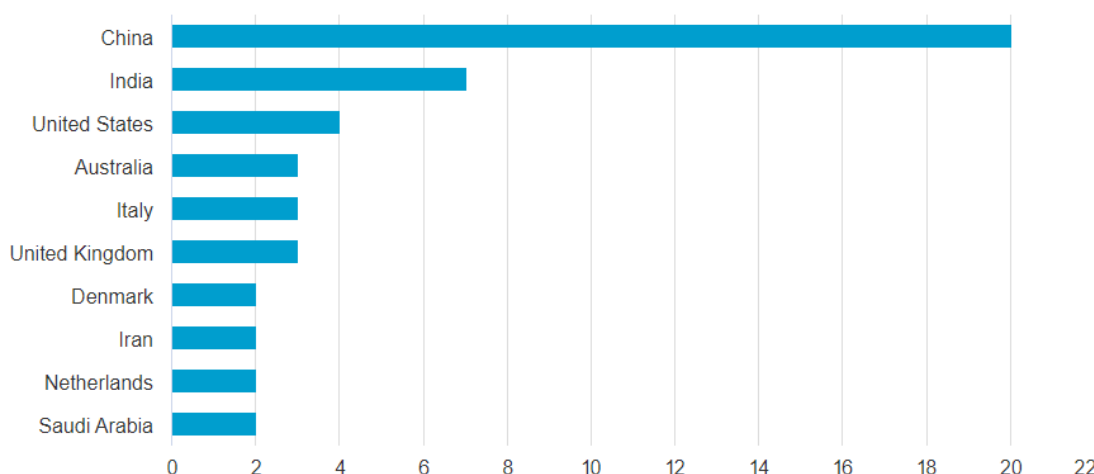
Em uma análise inicial dos artigos levantados nas bases, foi observado que o escopo desse trabalho atravessa não apenas diversas áreas de conhecimento, sendo produtores de literatura acadêmica sobre o tema áreas como Engenharias, Matemática, Ciências Sociais, Medicina, Agronomia e em sua maioria, da área de Ciência da Computação. Mas também observado que essa área de conhecimento é um problema compartilhado globalmente, sendo países como China, Índia e Estados Unidos os maiores produtores de artigos relativos a gestão e governança de dados.

FIGURA 8 - ARTIGOS POR ÁREA DE CONHECIMENTO



Fonte: ScienceDirect (2023).

FIGURA 9 - ARTIGOS POR PAÍS (SCIENCEDIRECT)



Fonte: ScienceDirect (2023).

4.1.1 TENDÊNCIAS EM GOVERNANÇA DESCENTRALIZADA: SAÚDE E IOT

A governança descentralizada se caracteriza por um modelo de governança onde o controle e responsabilização é distribuído na figura de múltiplas entidades, dentro do modelo descentralizado, o controle dos dados pode ser distribuído de muitas formas, uma movimentação e problemática atual é relativo a responsabilização e controle do armazenamento de dados pessoais, que pode possuir um modelo específico de controle, mas hoje depende do seu sistema controlador para guiar as melhores práticas.

Em se tratando de problemas de caracterização do controle e responsabilidade para com os dados pessoais, Martani et al. (2021) relata sobre o problema de *ownership* para com os dados de saúde especialmente no caso de países da União Europeia pela introdução de legislação geral de proteção aos dados pessoais, conhecida como *General Data Protection Regulation (GRPD)*, que foi um esforço regulatório que garantiu e se propôs a distribuir a responsabilização e direitos dentro do papel da governança de dados para o indivíduo ou sujeito produtor e serviço controlador dos dados pessoais.

Porém, segundo os autores, trouxe diversas lacunas legais, problemas de transparência e clareza sobre como a utilização dos dados pessoais especialmente na área da saúde pode ser realizada. Em um processo de entrevistas a especialistas legais, pesquisadores e outros atores da área da saúde, os autores encontraram muita

dificuldade e incerteza na caracterização do *data owner*, especialmente objetivando a proteção da privacidade de pacientes juntamente com a possibilidade de reutilização dos dados de saúde para o contínuo desenvolvimento de pesquisas na área, e abordam uma possibilidade de uma descentralização da posse e controle dos dados onde os pacientes realizariam a transferência da guarda, controle e processamento dos seus dados para utilização em pesquisas futuras (MARTANI et al., 2021).

Segundo Prieëlle, Reuver e Rezaei (2022) uma das características da internet das coisas (IoT) e seus dispositivos é a sua produção massiva de dados em velocidade, que quando utilizados em escala possuem um melhor aproveitamento. Sistemas diferentes necessitam de dados diferentes para melhorar seu serviço prestado, e ao mesmo tempo sistemas – especialmente no caso de sistemas IoT – adquirem dados que não necessariamente são ligados especificamente ao seu serviço prestado ou da sua organização, portanto o valor dos dados em escala é mais bem aproveitado quando compartilhados entre organizações, porém a adoção desse modelo de compartilhamento ainda é baixo justamente por problemas de controle e governança.

Em se tratando sobre o ecossistema de governança de dados em dispositivos IoT para a agricultura, e buscando identificar as maiores dificuldades na implementações desses sistemas para esse mercado, os pesquisadores encontraram um problema inesperado dentro da dificuldade de definição da figura do controlador e responsável pelos dados gerados dentro do sistema e processos relativos a produção horticultura, sendo identificados alguns fatores chaves para esses sistemas como a definição de *data owner* e acesso, critérios de determinação do *data owner* para dados gerados, e esse é considerado o fator mais importante, pois está diretamente relacionado a qualidade, acesso, controle e compartilhamento dos dados gerados (PRIEËLLE; REUVER; REZAEI, 2022).

Em sistemas onde o dado extraído é em essência um dado pessoal, a figura do *data owner* se confunde e não é clara, algumas vezes apontado na figura do usuário, outras na figura do sistema. Na literatura essa confusão permeia diversas pesquisas, e se reflete nas discussões de responsabilidade, no controle e gestão dos dados. Bo Yin, Hao Yin e Yulei Wu (2020), em seu artigo analisando a explosão da geração de dados de dispositivos IoT e da proliferação da internet das coisas, eles ponderam sobre esse problema, onde existe uma ambiguidade sobre a propriedade

dos dados e a necessidade de melhor estabelecimento de governança de dados nessa indústria.

Para os autores, a principal regra para a governança de dados de dispositivos IoT está na noção de que quem coleta o dado, esse portanto possui controle e responsabilidade sobre o dado coletado, se tornando de fato o *data owner*, portanto o sistema que coleta possui essa responsabilidade. Empresas digitais possuem portanto um grande volume de dados pessoais e a partir dessa noção se utilizam do processamento, uso e disseminação desses dados, e pela falta de alinhamento de mecanismos de governança desses tipos de dados, naturalmente são levadas a conflitos de privacidade, propriedade que podem se caracterizar em conflitos tanto éticos como legais, especialmente relativo ao compartilhamento entre diferentes organizações, sendo a privacidade um fator atual de preocupação (B. YIN; H. YIN; Y. WU, 2020).

Especialmente no caso de dispositivos IoT que produzem uma vasta quantidade de dados pessoais e em rápida velocidade, que exige uma maior transparência e segurança no controle e responsabilização, e um campo onde a governança de dados descentralizada encaixa por existir um inevitável compartilhamento de responsabilidade entre a organização provedora dos serviços e os usuários desses dispositivos.

Em seu artigo, Aslam, Bukovski e Mrissa (2021), ao construírem um modelo de gestão de dados especificamente para sistemas de energia elétrica descentralizadas ao se tratar de produção de energia comunitária renovável como eólica, solar fotovoltaica, dentre outras. O artigo trata especificamente de um modelo de controle dos dados com preocupação especificamente na privacidade de cada unidade produtora de energia para com seus dados, e seu controle de leitura e escrita na arquitetura de dados. Porém, isso traz também consigo um problema da governança desses dados, não no caráter técnico que é bem resolvido pelo seu modelo, mas sim no caráter do controle e responsabilização, que segundo os autores a posse imediata do controle dos dados dentro do modelo deve ser dada a um representante centralizador da comunidade, mas um problema inevitável descrito pelos próprios autores é que a responsabilização e controle é discutivelmente também um papel a ser compartilhado e descentralizado entre os membros da comunidade produtora dos seus dados.

Outro ponto presente nesse artigo, e também em demais artigos encontrados na literatura pertinente, é utilização de *blockchain* como ferramenta de descentralização de sistemas de armazenamento e controle de dados, instrumento que hoje é cada vez mais utilizado em larga escala para promoção da segurança e descentralização de sistemas, distribuindo não apenas a carga de processamento como também criando uma responsabilização sistemática para com o armazenamento e distribuição de dados pelos sistemas (ASLAM; BUKOVSKI; MARISSA, 2021).

4.1.2 GOVERNANÇA E *BLOCKCHAIN*

Ao se tratar de dados pessoais, sejam eles em dispositivos IoT, dados pessoais de saúde, dados de produção de energia, o problema é similar onde ocorre uma vasta e rápida produção de dados que são intrinsecamente pessoais, porém de responsabilização compartilhada. Nesse ponto, novamente a essência do problema é na pergunta de como descentralizar essa figura do *data owner* sujeito produtor e serviço controlador dos dados produzidos.

Para tanto, algumas alternativas surgem, especialmente no aspecto das ferramentas possíveis para que uma governança descentralizada aconteça da melhor forma possível, uma dessas ferramentas que consegue lidar com a abstração da descentralização da governança é a do *blockchain*.

Em seu artigo, Gao et al. (2021) tratando sobre uma solução em segurança e controles de acesso no compartilhamento de dados pessoais, os autores relatam o mesmo problema supracitado da privacidade dos dados, onde ao utilizar um serviço um indivíduo criador dos dados não possui mais absoluta posse, e o controle e responsabilização recai sobre a organização que armazena e processa esses dados, dentro dessa perspectiva os autores trazem a ferramenta *blockchain* como solução que consegue precisamente prover um estrutura que ao criptografar e distribuir o armazenamento dos dados de forma anonimizada (*hashs*), mantendo o controle para o indivíduo criador inicial por meio dos contratos inteligentes de uso dos dados, descentralizando portanto a governança de armazenamento e processamento desses dados.

Chen et al. (2020), também trazendo a tecnologia emergente de *blockchain* e contratos inteligentes como uma forma de possibilitar melhor governança de acesso

e uso dos dados pessoais de forma descentralizada, os autores portanto apresentam um modelo de *blockchains* em consórcio para descentralização do controle e acesso a dados pessoais, e trazem a ideia de consentimentos contínuos por parte do indivíduo criador dos dados, a partir de contratos inteligentes validando com a entidade controladora do seus dados a utilização, o controle e acesso dos seus dados pessoais descentralizados demonstrando que a ferramenta possui características que se demonstram importantes em uma realidade de necessidade inerente de descentralização da figura responsável pela governança de dados.

4.2 ANÁLISE DOS RESULTADOS: CONFERÊNCIA E TENDÊNCIAS DA ÁREA

Nessa seção serão analisados documentos de conferências e outros documentos relevantes promovendo o compartilhamento de conhecimento e casos de sucesso da área de gestão e governança de dados.

4.2.1 CONFERÊNCIA *ENTERPRISE DATA WORLD*

Enterprise Data World é uma conferência anual focada na gestão e governança de dados que inclui palestrantes de diversas áreas incluindo indústria, tecnologia, finanças e saúde. A conferência abrange uma variedade de tópicos relacionados a gestão de dados como a modelagem, arquitetura, qualidade, dentre outros. Em suas apresentações e networking foi possível observar as opiniões de especialistas do mercado de gestão e governança de dados para não somente as melhores práticas da área como também observar movimentações de tendências de pensamentos e tecnologias relevantes da área.

Em painel sobre tendências na arquitetura de dados, O'Brien, Durbank, Yuhasz (2022) destacam um impacto recente sobre a gestão e governança de dados dada uma mudança de mentalidade especialmente na importância da governança de dados para o contexto geral de gestão, assim como tecnologias de virtualização, *data mesh* e *data fabric*, que retiram a governança de um modelo centralizado para formas mais distribuídas e governadas, especialmente se utilizando de novas arquiteturas distribuídas que contribuem para essa mudança mais aprofundada na governança.

Um problema levantado e discutido sobre os efeitos da descentralização da governança de dados é justamente da reformulação de como ajustar os envolvidos nesse processo de controle e responsabilização de forma adequada sem prejudicar os processos de gestão de dados existentes nas organizações, principalmente por novos modelos terem uma abordagem de mudança de mentalidade das áreas de negócio, e naturalmente irá ocorrer uma interdependência entre a mudança estrutural e de arquitetura e a mudança cultural que a organização deve realizar para satisfazer os requisitos necessários para o correto desenvolvimento de modelos descentralizados de governança de dados, o que para algumas organizações pode ser um processo dificultoso. (O'BRIEN; DURBANK; YUHASZ, 2022)

Em uma nova tendência do mercado de arquitetura de dados, Crittenden (2022) discute sobre a inclusão de uma nova ferramenta de arquitetura e gestão de dados que seria uma forma de nível abstrato de compartilhamento e uso de dados dentro de organizações chamado de *data fabric*, especialmente para organizações que possuem soluções de arquiteturas de dados híbridas e desconexas, e existindo diversos tipos de soluções cloud como *datalake*, *datawarehouse*, e outros tipos de bancos de dados na nuvem e plataformas que são utilizadas em paralelo durante o ciclo de vida dos dados dentre sua produção até sua utilização por uma área de negócio, o *data fabric* portanto seria uma ferramenta utilizada para unificar esses sistemas de origem de dados e processos, e reduzindo os custos contínuos do gestão de dados.

No aspecto de governança de dados, a ferramenta de *data fabric* permite a criação de políticas e regulações que automaticamente são aplicadas aos diversos ambientes e sistemas de origem de dados, e possibilitando um processo de *compliance* mais consistente, efetivo e automatizado, e por consequência uma melhor confiabilidade, proteção, segurança dos dados. (CRITTENDEN, 2022)

Hirshhorn e Powell (2022), apresentando sobre os melhores processos em um modelo aplicado na prática para centralizado de governança de dados na organização Bank of New York Mellon, principalmente observando em um nível mais estratégico discutem sobre parcerias que são necessárias entre a governança de dados e as áreas de negócio, observando a importância de orçamentos, prazos, prioridades, ferramentas, juntamente com parcerias e conexão organizacional para redução de riscos e custos dentro de um processo de governança integrada de dados. Nessa análise de melhores práticas, o foco no orçamento próprio planejado e prazo alinhado

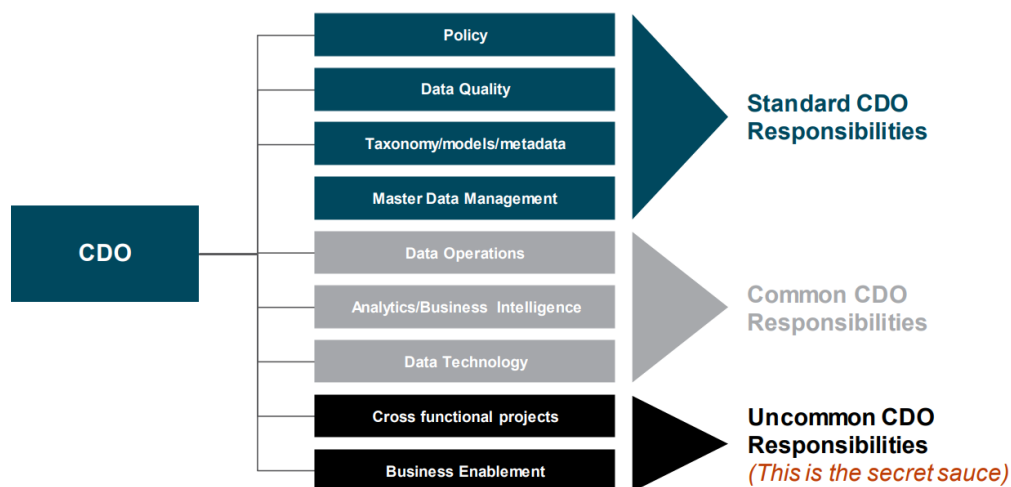
com os parceiros de negócio tem seu destaque por garantir uma responsabilização do próprio negócio para com o sucesso da governança de dados dentro do seu escopo. Para tanto, os alinhamentos com a gestão de tecnologia da organização são também peça chave para a garantia de interoperabilidade e gestão integrada dos dados dentro dos processos contínuos existentes e novos sistemas sendo criados pelas organizações.

Em se tratando de responsabilização da governança de dados, nessa organização é aplicado o modelo de gestão em uma área centralizada, chamada de Central Data Office (CDO), e são citadas as responsabilidades comuns dessa área como formulação de políticas, taxonomias, modelos, metadados, masterização e qualidade de dados. Mas também, de algumas responsabilidades comuns de serem adquiridas como as operações de dados, *analytics* e *business Intelligence* (BI) e tecnologias relacionadas com a gestão e governança de dados.

Sob destaque nessa análise de responsabilização, é citado o papel de realização juntamente com as áreas de negócio de projetos que não necessariamente tem como foco os dados, são uma importante oportunidade de demonstração de melhoria dos fluxos de dados para esses projetos, e também, um importante processo é o de possuir um time interno dentro do próprio CDO, com especialistas que possuam conhecimentos técnicos e de negócio, que irão realizar maior contato com a área de negócio, e que auxiliem os projetos a aplicarem da melhor forma possível as políticas e práticas da governança de dados alinhada a estratégica de cada área específica a ser atendida.

FIGURA 10 - ORGANIZAÇÃO ESTRUTURAL

Embed Partnership in the Org structure



FONTE: Hirshhorn e Powell (2022)

Nessa estrutura, em um contraponto a governança federada discutem a importância de implementações de áreas como metadados e qualidade de dados, que especialmente se controlados de forma centralizada por sua preocupação com generalização conseguem garantir uma melhor obtenção de valor dos dados, principalmente em áreas críticas para a organização, e observam também a necessidade de mudança de mentalidade onde os dados não fazem parte de um processo de negócio, e portanto a quebra dos silos e parcerias com áreas de negócios são essenciais para que a gestão e governança de dados seja bem sucedida (HIRSHHORN; POWELL, 2022).

Sequeda (2022), em apresentação sobre tendência especificamente para a área de governança de dados em *Big Data Analytics* (BDA) discute sobre o modelo *data mesh* dentro de uma perspectiva da necessidade de mudança da forma de se pensar na governança de dados no futuro. Pelo aumento da quantidade, variedade, velocidade em escala que os dados estão tomando nas organizações, existe a necessidade da escala da gestão e governança por consequência, e para tal a manutenção de modelos monolíticos de gestão e governança de dados não são mais suficientes para as dificuldades atuais, e criando gargalos de difícil resolução nas áreas de tecnologia e negócio, principalmente se tratando de responsabilização e controle dos dados. Portanto, uma mudança de certa forma estrutural, mas principalmente cultural, onde para a realização de uma gestão de dados em

organizações apropriada em escala os membros da organização devem observar o valor dos dados não como subprodutos de processos que realizam, mas sim como produtos de valor por si só.

4.2.2 GOVERNANÇA FEDERADA

Visto a movimentação de transformação das organizações, que passam pela conjuntura identificada por esse trabalho, e, portanto, necessitam de uma abordagem eficiente e mudança estrutural que promova a orientação da tomada de decisão para com os seus dados. Principalmente na última década com a adoção da nuvem como fonte primordial do armazenamento de ativos digitais, e crescimento da área de *Big Data Analytics (BDA)*, em uma movimentação muito rápida diversas formas de arquitetura foram sendo testadas, como o *data lakes*, *data warehouses*, dentre outras formas de armazenamento de dados dado seu volume e velocidade. Com essa nova realidade, a dificuldade da utilização de métodos de gestão e governança centralizada de dados causam justamente o problema do controle, responsabilização na figura do *data ownership*, sendo, portanto, uma falha de governança, e tendo como consequência diversos problemas nos aspectos da gestão de dados como a qualidade, a integração e interoperabilidade, a masterização, os metadados, dentre outros.

Algumas novas abordagens surgem justamente na área de BDA para a gestão e governança desse tipo de dados especificamente em organizações que investem nessa área como um dos pilares de sua tomada de decisão. Em seu artigo, Machado, Costa e Santos (2022) discutem sobre novas abordagens e novos paradigmas de arquitetura e governança de dados, e trazem o *data mesh* como um novo paradigma para essa governança por meio da descentralização e democratização do *data owner* dentro dos seus domínios. Baseado em casos de aplicações atuais em organizações como Netflix e Zalando, que se utilizaram dessa metodologia como estratégia organizacional e reformularam as estruturas de governança existentes dentro desse novo contexto. Alguns problemas elencados pelos autores na organização Netflix são a duplicação de esforços para conectar os dados em diferentes times, o peso e esforço na criação dessas conexões, e a dificuldade de implementações de boas práticas nos processos dentro do ciclo de vida dos dados que possuem em essência alto volume e velocidade.

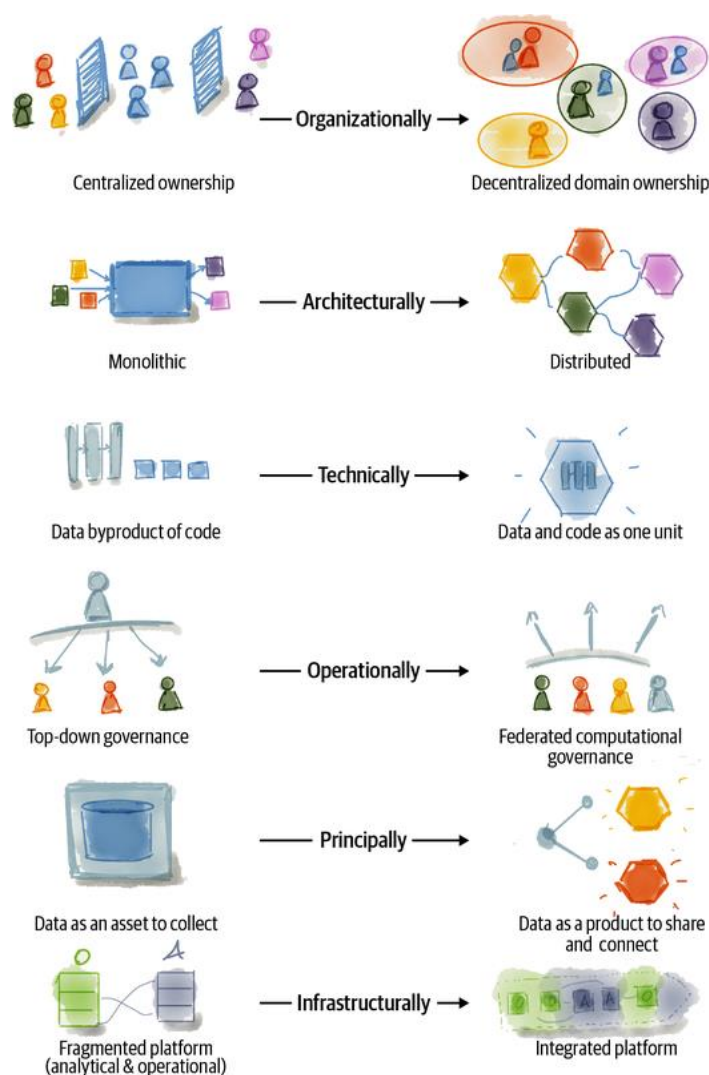
Ambas os casos dessas organizações buscando como prioridade a redução de sua complexidade operacional e gargalos, especificamente em seus processos de armazenamento, transformação e análise dos seus dados tiveram como resultados elementos técnicos e culturais que permitiram superar essas dificuldades operacionais, portanto a aplicação do *data mesh* não é sobre uma tecnologia específica a ser aplicada em uma arquitetura, mas sim fazer com que o problema dos dados seja uma preocupação organizacional que exige uma mudança estrutural para a melhor composição de uma governança apropriada.

4.2.3 ANÁLISE SOBRE A GOVERNANÇA FEDERADA: DATA MESH

Segundo Dehghani (2022), *data mesh* é uma abordagem sociotécnica descentralizada no processo de compartilhamento, acesso e gestão de dados analíticos (BDA) em ambientes complexos e de larga escala internos e externos a organizações. Para a autora, esse modelo propõe uma mudança de premissas, arquiteturas, soluções técnicas e sociais dentro de organizações sobre como a gestão de dados deve ser realizada ao se tratar de dados em escala.

Interessante observar que a mudança proposta por Dehghani (2022) se baseia não somente na arquitetura dos dados, mas sim em toda a filosofia da organização com relação aos seus dados, assim como as estruturas que permeiam o ciclo de vida dos dados existentes nas organizações. No aspecto organizacional, tecnológico, operações, filosofia, infraestrutura e arquitetura dos dados. Mudança elencada na descentralização do controle dos dados para os domínios de negócio que possuem maior proximidade, assim como distribuição da arquitetura dos dados, a federação da sua governança, o compartilhamento dos dados como produto a ser conectado com outros produtos e serviços e integração de plataformas de operações a análise.

FIGURA 11 - MUDANÇAS ORGANIZACIONAIS DATA MESH



FONTE: DEGHANI (2022, p.5)

O modelo *data mesh* possui quatro princípios relacionados ao processo de aumento de valor dos dados em escala, a sustentação da agilidade dos processos no crescimento das organizações e seus dados por consequência, assim como a complexidade e volatilidade dos contextos organizacionais. No quadro a seguir são representados os quatro princípios do *data mesh* juntamente com uma essência de sua aplicação na realidade.

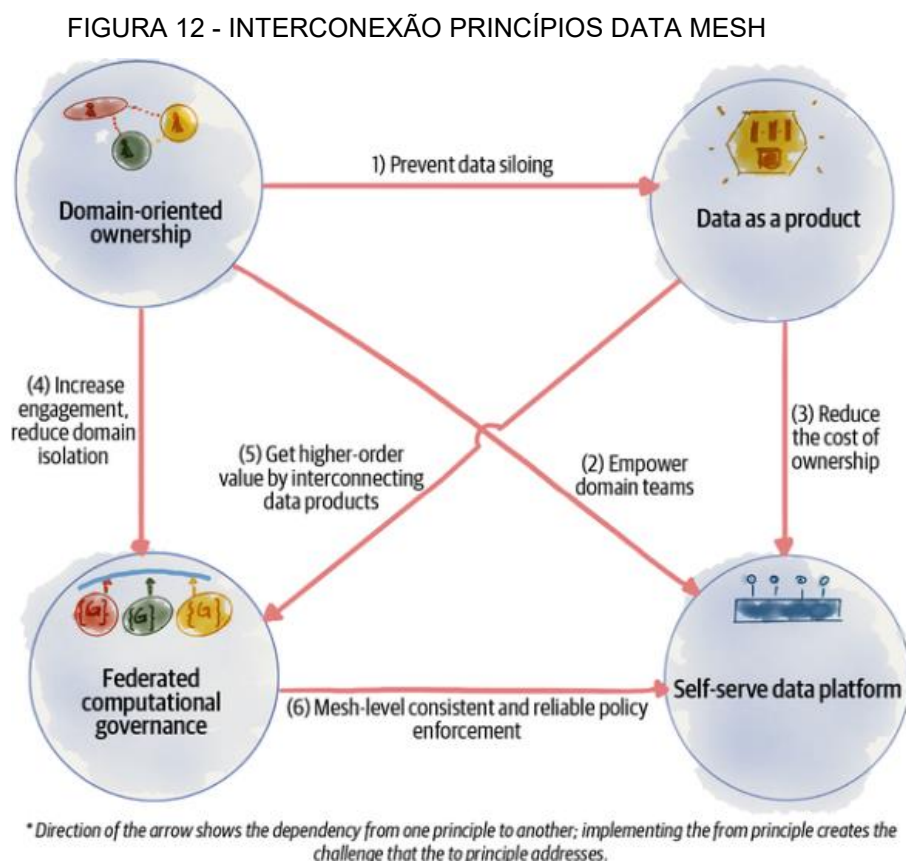
QUADRO 2 – PRINCÍPIOS DO MODELO *DATA MESH*

Princípios	Descrições
Controle do Domínio	Princípio de descentralização do <i>data owner</i> , onde o domínio de negócio que

	<p>está mais próximo relativo ao ciclo de vida do dado em específico (fonte criadora ou principal consumidor) recebe imediatamente seu controle e responsabilização para com esses dados.</p>
Dados como Produto	<p>Dados são tratados como um produto em si, específico de um domínio – ideia que vem em oposição a visão de serem um subproduto, resultados ou resíduos de um processo de negócio. Esses dados como produto devem ser compartilhados dentro das organizações, com seus os princípios próprios e autonomia dentro do seu contexto de ciclo de vida e modelos.</p>
Plataformas com autosserviço	<p>Princípio relacionado a remoção da fricção do compartilhamento de dados nas organizações, plataformas <i>self-service</i>, e serviços que promovem essas plataformas para cada produto de dado devem permitir a descoberta, acesso e uso dos dados em um fluxo aberto e transparente dentro da organização durante seu ciclo de vida.</p>
Governança Federada	<p>A governança dos dados dos domínios deve ser realizada em uma estrutura descentralizada de responsabilização e tomada de decisão por meio dos representantes do domínio, plataformas e áreas de negócio.</p>

FONTE: Dehghani (2022).

Esses princípios se interconectam dentro do contexto do que eles buscam corrigir, a prevenção de silos de dados, a redução de custo da responsabilização para com os dados dos domínios, o empoderamento dos domínios para com seus dados, o aumento da participação dos domínios e redução do isolamento desses na tomada de decisão para com os seus dados no ciclo de vida do BDA, assim como um aumento na consistência e confiabilidade dos dados. A figura a seguir tem como objetivo visualizar essa interconexão dos quatro princípios.



FONTE: DEGHANI (2022, p.9)

Dentro desses princípios, visto seu contexto, o aspecto da governança federada ganha uma importância visto que é a estrutura que controla o modelo e o sustenta em seus processos, uma governança federada com base no modelo do *data mesh* assegura não apenas a segurança, padronização, qualidade e interoperabilidade dos dados dentro dos seus domínios como também auxilia na comunicação dentro dos sistemas e processos distribuídos dentro do ciclo de vida dos dados.

Segundo Vijay Butte e Sujata Butte (2022) esse princípio da governança federada causa não apenas uma mudança na visão dos dados dentro das organizações como também uma mudança estrutural, onde lideranças estratégicas dessas organizações, assim como seus especialistas internos nos domínios, devem se preparar para essa nova filosofia. Portanto é interessante realização juntamente com um processo de treinamentos, recursos informacionais, feedback e quebra da implementação com iniciativas de pequenas descentralizações que movem progressivamente os dados da organização para o novo modelo.

O princípio da governança federada vem como um pilar que ajusta o problema causado pela descentralização que é justamente a interoperabilidade e padronização dos dados e da comunicação entre os membros desse processo de gestão e governança dos dados, portanto o equilíbrio da autonomia do domínio e a interoperabilidade global dos dados é uma preocupação que deve ser considerada por meio de um constante alinhamento. O modelo de governança federada possui portanto, um time de governança federada, que é composto por um comitê multidisciplinar e de responsabilidade compartilhada por especialistas de negócio, especialistas de segurança, *compliance* e jurídico, representantes dos domínios, especialistas das plataformas, facilitadores, gestão estratégica, dentre outros.

Na figura do *data steward*, dentro do modelo *data mesh*, existe o representante do domínio como figura chave, que possui proximidade local para tomada de decisão nos seus dados, assim como uma visão global dos problemas e decisões a nível organizacional por meio dos comitês de governança federada, ele portanto tem como contribuição a definição de políticas para a governança do seu dado como produto organizacional.

Esse processo tem como produto portanto, políticas locais apoiadas e suportadas em políticas globais de princípios, guias e regulamentos para realização da manutenção, acessibilidade, qualidade, modelagem, e outras formas de gestão dos dados a nível local que também se refletem uma interoperabilidade global dos dados como produtos dos domínios (DEGHANI, 2022).

4.3 DISCUSSÃO DOS RESULTADOS

A partir dos resultados encontrados, selecionados e analisados foram observados alguns padrões conceituais que podem auxiliar na responsabilização correta do *data owner*. Foi observado que, independentemente se os dados que passam por um processo de gestão, são de natureza dados pessoais ou organizacionais, a dificuldade persiste de identificação do responsável por existir em essência uma coleta massiva de dados analíticos, que necessitam de princípios de gestão de dados para que sejam bem sucedidas, e por consequência uma governança de dados efetiva, que tem como uma das suas principais figuras a do *data owner*.

A governança de dados é retratada no geral como um esforço multifuncional de processos colaborativos dentre diferentes áreas dentro da organização, que realiza o equilíbrio da acessibilidade, proteção e controle dos dados por meio de ferramentas, tecnologias, metodologias e processos e que asseguram em organizações por meio do exercício de autoridade e controle a melhor extração de valor dos dados para uma boa tomada de decisão centrada nos dados.

Um aspecto importante que diferencia a visão da gestão e governança de dados da literatura acadêmica em comparação com demais fontes práticas e de mercado, é o enfoque no aspecto do negócio como fonte primária de resolução de muitos conflitos principalmente no aspecto de responsabilização. Existem também aspectos de novos modelos que a literatura até o presente momento não permeou e que se demonstram importantes avanços no pensamento da gestão e governança de dados da atualidade.

Importante ressaltar que a partir da análise realizada, foi observado que a diferenciação da governança está relacionada a tipos diferentes de dados, dentro do contexto dos artigos analisados fica clara uma distinção das considerações para com a governança sobre dados transacionais, dados referenciais e dados de Big Data Analytics. E a governança podendo ser realizada de forma centralizada, descentralizada ou federada a depender do tipo de dado, estratégica organizacional, ciclo de vida e necessidade de negócio.

Alguns modelos de governança de dados, como o *data mesh*, são de uma visão e metodologia específica para com os dados organizacionais relacionados a BDA, visto que o formato de responsabilização federada é um complexo sistema que pode acarretar a desconexão do controle necessário no caso de dados críticos,

referenciais e mestres. Portanto, o ponto crucial para decisão da melhor estratégia de governança e por consequência da figura do *data owner* pode partir da análise do tipo de dado e do tipo de gestão necessária para com esse dado, podendo ocorrer diversas formas de governança dentro de uma área de negócio organizacional a depender da sua necessidade específica, projetos envolvidos, contexto e cultura organizacional.

Portanto, os modelos não se sobrepõem ou substituem, mas sim podem trabalhar em sintonia dentro da organização visto que tipos de dados e estruturas organizacionais diferentes demandam gestão e governança adaptada. Assim como, os diferentes modelos podem funcionar ou falhar a depender do alinhamento estratégico da organização e das unidades de negócio que produzem, que armazenam ou demandam dados para seu uso, visto que parcerias entre negócio e áreas de dados são fundamentais para a consolidação de qualquer proposta de governança.

5 CONSIDERAÇÕES FINAIS

A determinação do papel do *data owner* irá necessariamente variar a depender do alinhamento estratégico da organização, o tipo de dado como produto e ativo organizacional, a cultura organizacional, os alinhamentos sociotécnicos para com a gestão dos dados, dentre outros.

O *data owner* pode ser definido de forma centralizada, sendo uma figura responsável pelo dado produzido, armazenado, processado e utilizado, principalmente com relações a dados críticos sejam esses pessoais ou organizacionais, em uma única figura de autoridade e controle que pode ser uma pessoa ou um grupo de pessoas geralmente elencados para essa função em específico, e que portanto devem realizar um esforço a parte para sua conexão e parceria e alinhamento com as áreas de negócio, desde a priorização, orçamento, escopo, políticas, melhores práticas, adaptadas para que a governança seja bem sucedida.

Pode também ser definido de forma descentralizada, onde existe uma distribuição de sua responsabilização em diversas figuras ou entidades, em casos encontrados de necessidades atuais de tratamento de dados pessoais produzidos, armazenados e distribuídos por vezes em larga escala tendendo naturalmente a distribuir também sua responsabilização, onde o indivíduo produtor do dado não necessariamente irá administrar ou ter qualquer controle pelo armazenamento e processamento do seu dado pessoal, que é de sua propriedade. Portanto, em uma forma descentralizada, é essencial que os acordos entre as partes tenham a devida transparência, clareza e separação de responsabilidades elencadas entre as partes, seja por contratos formais diretos, contratos inteligentes, interações contínuas, os diferentes agentes dessa forma de governança devem estar alinhados para garantia da segurança, privacidade e controle dos dados.

Já em um formato federado de governança de dados, o principal achado nesse trabalho foi o modelo *data mesh*, que de certa forma possui características de ambas centralização e descentralização, geralmente se tratando de dados organizacionais que são produzidos, armazenados e processados em alta quantidade, variedade e velocidade, possui também naturalmente uma complexidade maior. Portanto, no caso da governança federada, e necessitando a quebra da figura do *data owner* como uma responsabilidade coletiva, e dentro da própria separação desses

dados em produtos dos seus domínios que são autônomos em sua tomada de decisão, os controladores e responsáveis para com a governança se aproximam o máximo possível da área afetada pelas próprias decisões de governança.

Ainda que devam existir as políticas de um aspecto global de interoperabilidade, a governança se aplica em um processo mais indutivo. A aplicação da responsabilidade e controle para com os dados dos produtos está na figura de comitês internos locais e globais formados pelos representantes dos produtos e domínios que controlam seus dados, que definem em conjunto as melhores práticas e políticas tanto na perspectiva global da estratégia organizacional, quanto na perspectiva local da estratégia das áreas de negócio.

O gestor da informação pode atuar dentro desse aspecto, principalmente ligado a inerente dificuldade observada nessa pesquisa da junção de conhecimentos de áreas de negócio alinhados com conhecimentos de tecnologia, visto que ambos devem andar em paralelo para que uma governança de dados ocorra de forma efetiva, é uma lacuna que pode ser área de atuação do profissional do gestor da informação que pode atuar dentro dos modelos apresentados para com a governança de dados atuando com ambos os conhecimentos e práticas das áreas de dados, tecnologia e administração de negócios.

Esse trabalho objetivou observar na literatura e na prática, a aplicabilidade dos estudos para com o papel do *data owner*, a evolução, a diferenciação e importância da gestão e governança de dados na sua figura de autoridade e controle, assim como o esclarecimento de algumas melhores práticas existentes sobre a escolha do papel e responsabilidades do *data owner* dentro das organizações.

Dentro do escopo dessa pesquisa, algumas especificidades ainda necessitam continuidade da discussão e de um aprofundamento desse tema, principalmente no aspecto das atividades específicas do *data owner* como indivíduo, grupo ou organização dentro dos processos ligados a uma arquitetura centralizada ou distribuída. Como a atuação desse controle e responsabilização se dá nos processos de qualidade de dados, arquitetura de dados, dentre outros. Durante a pesquisa também foram observados novas tendências de tecnologias futuras que poderão ser utilizadas como ferramentas de governança de dados, a aplicação dessas ferramentas práticas para facilitar a integração desses complexos, porém necessários processos de governança dentro das organizações. Assim como legislações atuais cada vez mais regulamentam a governança dos dados e obrigam organizações a se adaptar.

REFERÊNCIAS

ABELL, J.; et al. Big Data-Driven Manufacturing—Process-Monitoring-for-Quality Philosophy. **Journal of Manufacturing Science and Engineering**. v. 139. out. 2017. Disponível em: <https://doi.org/10.1115/1.4036833>. Acesso em: 30 abr. 2022.

ABRAHAM, R; SCHNEIDER, J.; BROCKE, J. Data governance: A conceptual framework, structured review, and research agenda. **International Journal of Information Management**. v. 49, p. 424-438, dez. 2019. Disponível em: <https://doi.org/10.1016/j.ijinfomgt.2019.07.008>. Acesso em: 03 de abril de 2022.

ASLAM, S.; BUKOVSKI, V.; MARISSA, M. Decentralized Data Management Privacy-Aware Framework for Positive Energy Districts. **MDPI Energies**. v. 14, 2021. Disponível em: <https://doi.org/10.3390/en14217018>. Acesso em: 07 fev. 2023.

AIKEN, P.; HARBOUR, T. **Data strategy and the enterprise data executive: Ensuring that business and IT are in synch in the post-big data era**. New Jersey: Technics Publications, 2017.

AIKEN, P.; BILLINGS, J. **Monetizing Data Management: Finding the Value in your Organization's Most Important**. New Jersey: Technics Publications, 2013.

ARASS, M.; SOUISSI, N. **Data lifecycle: from big data to SmartData**. 2018 IEEE 5th International Congress on Information Science and Technology, p. 80-87. 2018. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8596547>. Acesso em: 25 abr. 2022.

BENFELDT, O.; Persson, J.; Madsen, S. Data Governance as a Collective Action Problem. **Information System Frontiers**. v. 22, p. 229-313, abr. 2020. Disponível em: <https://link.springer.com/article/10.1007/s10796-019-09923-z>. Acesso em: 03 abr. 2022.

BEVIR, M. **Governance: A very short introduction**. Oxford: Oxford University Press, 2012.

BOOTH, W.; COLOMB, G.; WILLIAMS, J. **The Craft of Research**. Chicago: University of Chicago Press, 2008.

BITTERER, A.; NEWMAN, D. Organizing for Data Quality, **Gartner Research**, Stamford, 2007.

BROUS, P.; JANSSEN, M. Trusted Decision-Making: Data Governance for Creating Trust in Data Science Decision Outcomes. **Administrative Sciences**. v. 10, out. 2020. Disponível em: <https://www.mdpi.com/2076-3387/10/4/81>. Acesso em: 03 abr. 2022.

CARRETERO, A.; et al. MAMD 2.0: Environment for data quality processes implantation based on ISO 8000-6X and ISO/IEC 33000. **Computer Standards & Interfaces**. v. 54, p.139-151, nov. 2017. Disponível em: <https://doi.org/10.1016/j.csi.2016.11.008>. Acesso em: 03 abr. 2022.

CHEN, Y.; CHEN, S.; LIANG, J.; FEAGAN, L.; WEILI, H.; HUANG, S.; WANG, X. Decentralized data access control over consortium blockchains. **Information Systems**. V. 94, 2020. Disponível em: <https://doi.org/10.1016/j.is.2020.101590>. Acesso em: 16 fev. 2023.

CRITTENDEN, M. Create Data Fabric Instead of Data Silos. In: **Enterprise Data World Digital**, 26., 2022.

DAKI, H.; et al. Big Data management in smart grid: concepts, requirements and implementation. **Journal of Big Data**. v. 4, p. 1-19, abr. 2017. Disponível em: <https://doi.org/10.1186/s40537-017-0070-y>. Acesso em: 25 abr. 2022.

DAMA - Data Management Association International. **DAMA-DMBOK: Data Management Body of Knowledge**. 2^a. Ed. New Jersey: Technics Publications, 2017.

DATA GOVERNANCE INSTITUTE. **How to Use the DGI Data Governance Framework to Configure Your Program**. Disponível em: <https://datagovernance.com/wp->

[content/uploads/2020/07/wp how to use the dgi data governance framework.pdf](#).

Acesso em: 04 mai. 2022.

DAVENPORT, T. *Ecologia da Informação*. 2 ed. São Paulo: Futura, 1998.

DYCHÉ, J.; LEVY, E. **Customer data integration**: Reaching a single version of the truth. New Jersey: Wiley, 2008.

DEGHANI, Z. **Data Mesh: Delivering Data-Driven Value at Scale**. O'Reilly Media. 2022.

EVANS, N; PRICE, J. Barriers to the Effective Deployment of Information Assets: The Role of the Executive Manager. **Interdisciplinary Journal of Information, Knowledge, and Management**. v. 7, p. 162-169. 2012. Disponível em: <https://doi.org/10.28945/1721>. Acesso em: 25 abr. 2022.

FISHER, T. *The Data Asset: How Smart Companies Govern Their Data for Business Success*. New Jersey: John Wiley & Sons, 2009.

GAO, H.; MA, Z.; LUO, S.; XU, Y.; WU, Z. BSSPD: A Blockchain-Based Security Sharing Scheme for Personal Data with Fine-Grained Access Control. **Wireless Communications and Mobile Computing**, 2021. Disponível em: <https://doi.org/10.1155/2021/6658920>. Acesso em: 07 fev. 2023.

GANTZ, J.; REINSEL, D. Extracting value from chaos. **IDC iview**. p. 1-12, jun. 2011. Disponível em: <http://www.kushima.org/wp-content/uploads/2013/05/DigitalUniverse2011.pdf>. Acesso em: 25 abr. 2022.

HIRSCHHORN, E.; POWELL, L. Data Governance's Secret Sauce – Partnership with Business. In: **Enterprise Data World Digital**, 26., 2022.

JANSSEN, M; et al. Data governance: Organizing data for trustworthy Artificial Intelligence. **Government Information Quarterly**. v. 37, jul. 2020. Disponível em: <https://doi.org/10.1016/j.giq.2020.101493>. Acesso em: 03 abr. 2022.

KATRI, V.; BROWN, C. Designing data governance. **Communications of the ACM**. v. 53, p 148-152, jan. 2010. Disponível em: <https://doi.org/10.1145/1629175.1629210>. Acesso em: 04 mai. 2022.

LADLEY, J. Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program. 1ª Ed. Burlington: Morgan Kaufmann, 2012.

LANEY, D. **3D data management: Controlling data volume, velocity and variety**. META Group Research Note. 2001. Disponível em: <https://studylib.net/doc/8647594/3d-data-management--controlling-data-volume--velocity--an>. Acesso em: 25 abr. 2022.

MACHADO, I.; COSTA, C.; SANTOS, M. Data Mesh: Concepts and Principles of a Paradigm Shift in Data Architectures. **Procedia Computer Science**. v. 196, p. 263-271, 2022. Disponível em: <https://doi.org/10.1016/j.procs.2021.12.013>. Acesso em: 16 fev. 2023.

MARTANI, A.; GENEVIÉVE, L.; ELGER, B.; WANGMO, T. 'It's not something you can take in your hands'. Swiss experts' perspectives on health data ownership: an interview based study. **BMJ Open**. 2021.

MCGEE, J.; PRUSAK, L. **Gerenciamento estratégico da informação**: aumente a competitividade e a eficiência de sua empresa utilizando a informação como ferramenta estratégica. Rio de Janeiro: Campus, 1994.

MCGILVRAY, D. Data governance: a necessity in an integrated information world. **DM Review**. v. 16, p. 24. dez. 2006. Disponível em: <https://www.proquest.com/docview/214674184?pq-origsite=gscholar&fromopenview=true>. Acesso em: 30 abr. 2022.

NASCIO. Data Governance – Managing Information as an Enterprise Asset: Part I – An Introduction. NASCIO Data Governance Series. 2008. Disponível em:

<https://www.nascio.org/wp-content/uploads/2019/11/NASCIO-DataGovernance-Part1.pdf>. Acesso em: 04 mai. 2022.

O'BRIEN, J.; BURBANK, D.; YUHASZ, G. Trends in Data Architecture – A Holistics View of Data Fabrics, Meshes, Lakes and Warehouses. In: **Enterprise Data World Digital**, 26., 2022.

PRIEËLLE, F.; REUVER, M.; REZAEI, J. The Role of Ecosystem Data Governance in Adoption of Data Platforms by Internet-of-Things Data Providers: Case of Dutch Horticulture Industry. **IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT**. v. 69, no. 4, p. 940-950, 2022.

REDMAN, T. *Data Driven: Profiting from Your Most Important Business Asset* Redman. Boston: Harvard Business Review Press, 2008.

REN, S.; et al. A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: A framework, challenges and future research directions. **Journal of Cleaner Production**. v. 210, p. 1343-1365, nov. 2019. Disponível em: <https://doi.org/10.1016/j.jclepro.2018.11.025>. Acesso em: 30 abr. 2022.

SANTOS, Paula Xavier; ALMEIDA, Bethânia de Araújo; HENNING, Patricia (Coord.). **Livro verde ciência aberta e dados abertos**: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional. Rio de Janeiro: Fiocruz, 2017. 141 p. Disponível em: <https://www.arca.fiocruz.br/bitstream/icict/24117/2/Livro-Verde-07-06-2018.pdf>. Acesso em: 16 abr. 2022.

SEINER, R. **Non-invasive data governance**: The path of least resistance and greatest success. New Jersey: Technics Publications, 2014.

SEQUEDA, J. How to Be Agile in a Data Mesh World. In: **Enterprise Data World Digital**, 26., 2022.

SESI. Departamento Regional do Paraná. Habilidades para a indústria 4.0: visão sistêmica. In: SESI. Departamento Regional do Paraná. **Skills 4.0**: habilidades para a indústria. Curitiba, PR: SESI/PR, 2020. Cap. 5. Disponível em:

https://longevidade.ind.br/wp-content/uploads/2020/02/livro_skills4.0.pdf. Acesso em: 26 abr. 2022.

STEVENS, J.; CARALLI, R.; WILLKE, B. **Information Asset Profiling**. Software Engineering Institute: Carnegie-Mellon University. jun. 2005. Disponível em: <https://apps.dtic.mil/sti/citations/ADA441305>. Acesso em: 25 abr. 2022.

VIJAY, B.; BUTTE, S. Enterprise Data Strategy: A Decentralized Data Mesh Approach. **2022 International Conference on Data Analytics for Business and Industry (ICDABI)**. Sakhir, Bahrain. 2022. Disponível em: <https://doi.org/10.1109/ICDABI56818.2022.10041672>. Acesso em: 16 fev. 2023.

VISCUSI, G.; BATINI, C. Digital information asset evaluation: characteristics and dimensions. *In*: CAPORARELLO, L.; DI MARTINO, B.; MARTINEZ, M. **Smart Organizations and Smart Artifacts**. Lecture Notes in Information Systems and Organization, 7^a Ed. Springer, mai. 2014. Disponível em: https://doi.org/10.1007/978-3-319-07040-7_9. Acesso em: 25 abr. 2022.

WEBER, K.; OTTO, B.; ÖSTERLE, H. One Size Does Not Fit All---A Contingency Approach to Data Governance. **Journal of Data and Information Quality**. v. 1, p 1-27, jun. 2009. Disponível em: <https://doi.org/10.1145/1515693.1515696>. Acesso em: 04 mai. 2022.

WILKINSON, M; et al. The FAIR Guiding Principles for scientific data management and stewardship. **Sci. Data**. mar. 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 25 abr. 2022.

YEARLEY, S. **Science, Technology, and Social Change**. Londres: Routledge, 2015.

YIN, B.; YIN, H.; WU, Y. FDC: A Secure Federated Deep Learning Mechanism for Data Collaborations in the Internet of Things. **IEEE INTERNET OF THINGS JOURNAL**. v. 7, n. 7, p. 6348-6359, 2020.