

ARQUITETURAS DESCENTRALIZADAS DE DADOS: DATA MESH, UMA REVISÃO SISTEMÁTICA

Decentralized data architectures: Data Mesh, a systematic review

Douglas Correia Rosseto
Graduando
Engenharia da Computação
douglascorreia@ucl.br

Marlon Ferrari
Orientador
Faculdade UCL
marlonferrari@ucl.br

A lista completa com informações de autoria está no final do artigo

RESUMO

Objetivo: Apresentar uma revisão e levantar discussões sobre o *Data Mesh*, uma arquitetura descentralizada de dados, abordando temas como validade, utilidade e aplicabilidade nas empresas modernas.

Método: Utilizando o método de revisão sistemática, foram selecionados artigos, pesquisas e casos de uso que atendessem aos critérios estabelecidos para a análise.

Resultado: Os estudos revisados revelaram uma crescente discussão em torno do conceito de *Data Mesh*, que visa superar as limitações das arquiteturas monolíticas existentes. Os resultados geralmente apresentaram uma visão positiva dessa abordagem. Um ponto comum nas discussões é o empoderamento dos produtores e consumidores de dados, que permite uma maior proximidade entre a área de negócios e os produtos de dados, resultando em entregas mais ágeis. A descentralização do gerenciamento de dados e a quebra de silos também foram destacadas como aspectos benéficos. No entanto, é importante ressaltar que ainda há lacunas de conhecimento em relação às limitações e orientações práticas para a implementação bem-sucedida do *Data Mesh*. Poucos estudos abordaram esses aspectos específicos, destacando a necessidade de pesquisas futuras para uma compreensão mais aprofundada das desvantagens e desafios associados à implementação do *Data Mesh*.

PALAVRAS-CHAVE: Arquitetura, *Data Mesh*, Monolítica, Descentralização; Revisão sistemática.

ABSTRACT

Objective: This paper aims to present a systematic review and raise discussions about Data Mesh, a decentralized data architecture, addressing topics such as validity, usefulness, and applicability in modern enterprises.

Methods: Using the systematic review method, articles, research papers, and use cases that met the established criteria for analysis were selected.

Results: The reviewed studies revealed a growing discussion around the concept of Data Mesh, which aims to overcome the limitations of existing monolithic architectures. The results generally presented a positive view of this approach. A common point highlighted in the discussions is the empowerment of data producers and consumers, enabling closer alignment between the business domain and data products, resulting in more agile deliveries. Decentralized data management and the breaking of silos were also highlighted as beneficial aspects. However, it is important to note that there are still knowledge gaps regarding the limitations and practical guidelines for successful implementation of Data Mesh. Few studies addressed these specific aspects, emphasizing the need for future research to gain a deeper understanding of the disadvantages and challenges associated with Data Mesh implementation.

KEYWORDS: Architecture, Data Mesh, Monolithic, Decentralization, Systematic review.

1 INTRODUÇÃO

Ao se tornar a força motriz da digitalização, os dados são considerados como um ativo-chave central por muitas empresas. Usados em uma escala cada vez maior na tomada de decisões e aplicações, há uma demanda constante por mais dados, de melhor qualidade e disponibilidade, e em um horizonte de tempo mais amplo (LOUKIALA, 2021). Diante deste desafio, soluções tecnológicas foram introduzidas, como *data warehouses* e *data lakes*, com o objetivo de integrar diversas fontes de dados em um repositório central e monolítico, que, ao longo do tempo, demonstraram suas limitações em termos de escalabilidade e custo. Pretende-se que a engenharia de dados siga o caminho da engenharia de software e abandone arquiteturas monolíticas, criando assim arquiteturas de dados descentralizadas (MACHADO; COSTA; SANTOS, 2021).

Dentro deste contexto surge o *Data Mesh*, uma abordagem sociotécnica descentralizada para gerenciar dados analíticos em escala que consiste na implementação de uma arquitetura onde os dados são distribuídos intencionalmente entre vários nós (*mesh*), de forma que não haja caos ou silos de dados, uma vez que existem estratégias de governança de dados centralizadas garantindo a distribuição por toda a organização (DONCEVI et al., 2022).

Este artigo apresenta uma revisão sistemática da literatura que possui como objetivo levantar estudos acerca do tema e identificar características deles, de forma a contribuir com a definição da arquitetura *Data Mesh*, vantagens, benefícios e como essa mudança de paradigma afeta a governança e privacidade dos dados.

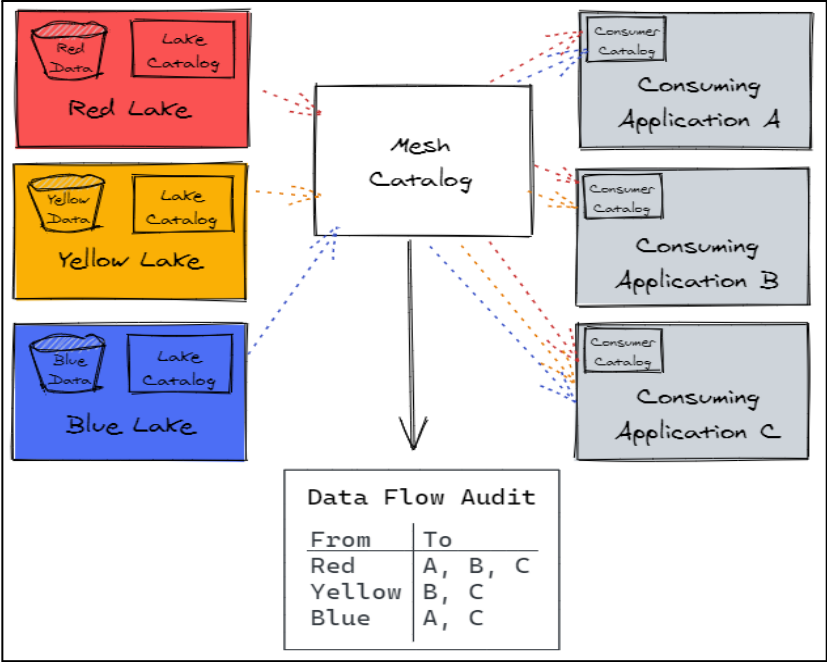
2 ARQUITETURA DATA MESH

Introduzido inicialmente pela Zhamak Dehghani, o *Data Mesh* tem como objetivo retirar conceitos de arquiteturas modernas distribuídas e aplicar nas arquiteturas de dados centralizadas, dentre eles estão: considerar os domínios como prioridade, aplicar a mentalidade de plataforma infraestrutura self-service de dados e tratar os dados como produto, a fim de reduzir a dependência de equipes de dados centralizadas e aumentar a colaboração entre as equipes (DEGHANI, 2020).

Para que o *Data Mesh* alcance seu propósito, ele deve se basear em quatro conceitos principais. O primeiro está relacionado à forma como os dados são organizados de acordo com a natureza da organização - propriedade e arquitetura de dados descentralizados orientados por domínio. O segundo conceito implica mudar a maneira como os dados são vistos dentro da organização - o conceito de dados como produto. O terceiro conceito está relacionado à transformação da infraestrutura de serviços em uma plataforma de dados de autoatendimento. Por fim, o conceito que evita o caos na Mesh - governança computacional federada (MACHADO; COSTA; SANTOS, 2021).

A Figura 1 abaixo retrata essa mudança de paradigma, nela é possível observar os diferentes *lakes* (*Red*, *Yellow* e *Blue*) como produtos separados pela organização, com a responsabilidade de produzir, armazenar e catalogar a ponto de deixar disponível e padronizado organizacionalmente no catálogo Mesh e, em seguida, ser distribuído aos consumidores, sejam aplicações, plataformas de visualização de dados, analistas de negócios e entre outros.

Figura 1: Exemplo de arquitetura simplificada *Data Mesh*.



Fonte: MORGAN s.d.

3 METODOLOGIA

Uma revisão sistemática, é uma forma de pesquisa que utiliza como fonte de dados a literatura sobre determinado tema. Esse tipo de investigação disponibiliza um resumo das evidências relacionadas a uma estratégia de intervenção específica, mediante a aplicação de métodos explícitos e sistematizados de busca, apreciação crítica e síntese da informação selecionada. (SAMPAIO; MANCINI, 2006).

3.1 IDENTIFICAÇÃO DO TEMA E SELEÇÃO DE PESQUISA

O presente estudo tem como base o seguinte objetivo de investigação sendo ele: levantar estudos acerca do tema e identificar características deles, de forma a contribuir para a definição da arquitetura *Data Mesh* e alinhar seus principais pilares.

Assim, a pesquisa bibliográfica partiu das seguintes questões:

- O que é a arquitetura *Data Mesh* e quais suas características?
- Quais as vantagens e benefícios da arquitetura?
- Como a abordagem do *Data Mesh* afeta a governança de dados e a privacidade dos dados em uma organização?
- Quais os principais desafios na estrutura organizacional?

3.2 CRITÉRIOS DE INCLUSÃO E EXCLUSÃO

Os artigos foram selecionados na base de dados *Google Scholar* no período entre janeiro e maio de 2023. Para realização das buscas foram usadas combinações de palavras-chave com o objetivo de estreitar os resultados, tais como: *data mesh*; *decentralized*; *architecture*; *governance*; *domain*, como os seus correspondentes em português e os seus devidos cruzamentos. Desta pesquisa, foram identificados 110 artigos.

No quadro abaixo, é apresentada a estratégia da pesquisa, em ordem crescente, utilizando a ferramenta de pesquisas avançadas do *Google Scholar* com as combinações de palavras-chave:

Quadro 1 - Resultados obtidos a partir da busca com palavras-chave em inglês

#	Palavras-chave	Resultados obtidos
Ordem	Base de dados:	Google Scholar
1	"Data Mesh"	1700
2	architecture "Data Mesh"	785
3	architecture decentralized "Data Mesh"	239
4	architecture decentralized governance "Data Mesh"	151
5	architecture decentralized governance domain "Data Mesh"	110
Total:		110

Fonte: o autor (2023).

Quadro 2 - Resultados obtidos a partir da busca com palavras-chave em português.

#	Palavras-chave	Resultados obtidos
Ordem	Base de dados:	Google Scholar
1	arquitetura "Data Mesh"	6
2	arquitetura descentralizada "Data Mesh"	5
3	arquitetura descentralizada governança "Data Mesh"	1
4	arquitetura descentralizada governança domínio "Data Mesh"	0
Total:		0

Fonte: o autor (2023).

Nesta etapa procedeu-se à seleção dos critérios de elegibilidade. Os critérios de inclusão e exclusão definidos encontram-se descritos no Quadro 3.

Quadro 3 - Critérios de inclusão e exclusão.

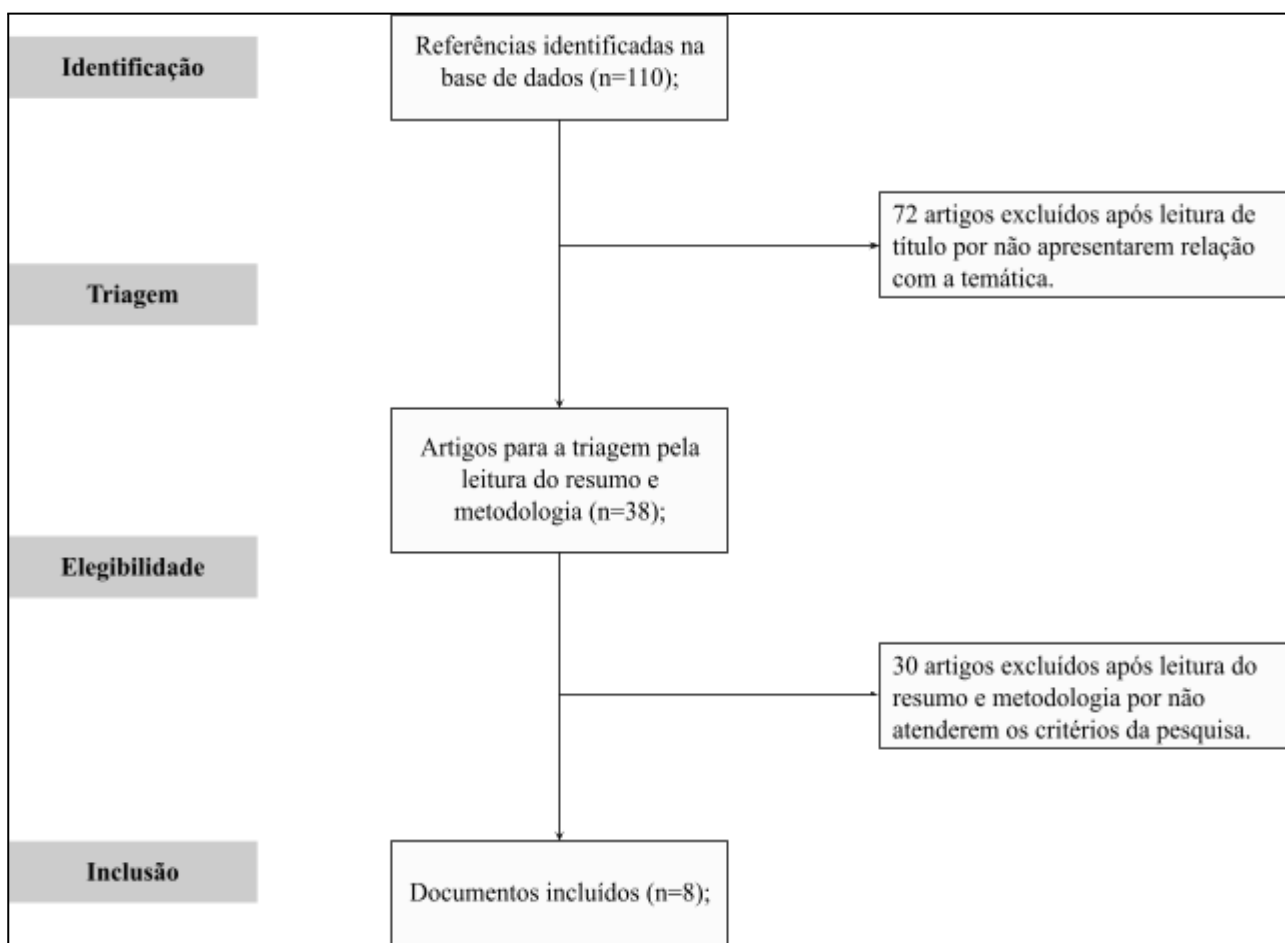
Critérios de Inclusão
Artigos publicados no período de 2019 a 2023;
Artigos publicados em português e inglês;
Artigos que abordem o <i>Data Mesh</i> e/ou alguma das questões propostas;
Artigos e documentos disponíveis para leitura na íntegra.
Critérios de Exclusão
Artigos anteriores à data de publicação (2019 - 2023);
Artigos e documentos em outras línguas que não as incluídas;
Artigos não disponíveis na íntegra;
Artigos de revisão de todos os tipos, teses e dissertações e cartas ao editor.

Fonte: o autor (2023).

4 RESULTADOS

Nesta etapa foi realizada a análise dos títulos, resumos e palavras-chave de todas as publicações identificadas e verificamos os critérios de elegibilidade estabelecidos inicialmente. Na Figura 2, podemos, então, observar o diagrama da presente revisão sistemática, desde a identificação dos artigos até à sua inclusão. Recorreu-se ao modelo PRISMA para se descrever o fluxo de informações através das diferentes fases da revisão, mapeando o número de registros identificados, incluídos e excluídos e os motivos das exclusões.

Figura 2: Fluxograma da revisão sistemática.



Fonte: o autor (2023).

4.1 CATEGORIZAÇÃO DOS ESTUDOS SELECIONADOS

Para elencar as características necessárias ao projeto, foi utilizada uma matriz de síntese de forma a sumarizar e organizar as informações obtidas nas fases anteriores, representada no Quadro 5, com os artigos obtidos na seleção:

Quadro 5 - Caracterização dos artigos incluídos na revisão sistemática.

#	Ano	Autores	Título	Idioma	País
1	2021	Divya Joshi, Sheetal Pratik e Madhu Podila	Data governance in data mesh infrastructures: the Saxo Bank case study	Inglês	Índia
2	2021	Loukiala et al.	Migrating from a centralized data warehouse to a decentralized data platform architecture	Inglês	Finlândia
3	2021	Michał Zasadzinski, Michael Theodoulou, Markus Thurner e Kshitij Ranganath	The Trip to The Enterprise Gourmet Data Product Marketplace through a Self-service Data Platform	Inglês	EUA, Espanha e Suíça
4	2022	Y. Hooshmand, J. Resch, P. Wischniewski e P. Patil	From a Monolithic PLM Landscape to a Federated Domain and Data Mesh	Inglês	Alemanha
5	2022	Ismail Olaniyi MURAINA e Moses Adeolu AGOI	Which Enterprise Data Management Strategy Is Right For Data Scientists And Business Analysts?	Inglês	Nigéria
6	2022	Juraj Doncevič, Kresimir Fertilj, Mario Brcić e Mihael Kovac	Mask-Mediator-Wrapper architecture as a Data Mesh driver	Inglês	Croácia
7	2023	Jan Bode, Niklas Kühl, Dominik Kreuzberger, Sebastian Hirschl, Carsten Holtmann	Data Mesh: Best Practices to Avoid the Data Mess	Inglês	Alemanha
8	2023	Worapol Alex Pongpech	A Distributed Data Mesh Paradigm for an Event-based Smart Communities Monitoring Product	Inglês	Tailândia

4.2 ANÁLISE E INTERPRETAÇÃO DOS RESULTADOS

Nesta etapa, analisaram-se todos os dados obtidos com o intuito de retirar conclusões e, conseqüentemente, construir uma perspectiva crítica sobre os artigos analisados. Esta análise foi realizada, tendo por base a matriz de síntese elaborada na etapa anterior.

Após a aplicação dos critérios de inclusão e exclusão foram, então, selecionados 8 artigos, sendo todos escritos na língua inglesa e oriundos de diversos países. No Quadro 6 é possível identificar a origem dos artigos por zona do globo.

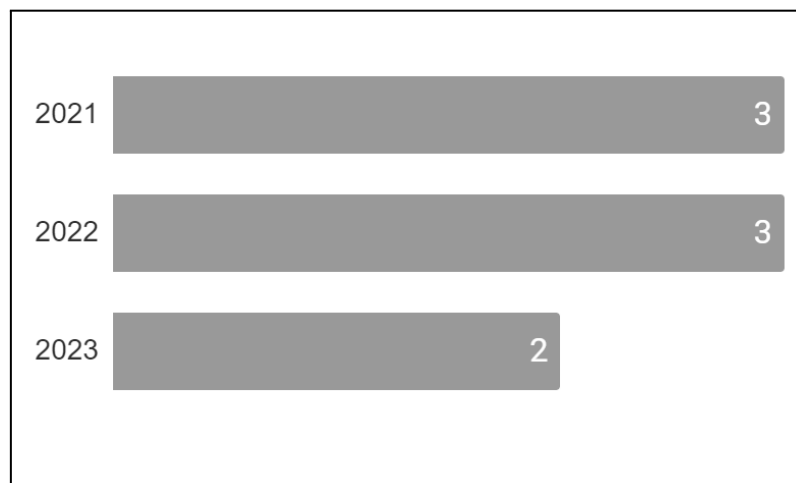
Quadro 6 - Quantidade de artigos por região.

Região	Nº de Artigos	Países
Europa	4	Alemanha, Finlândia, Espanha, Suíça e Croácia
América	1	EUA
Ásia	2	Índia e Tailândia
África	1	Nigéria

Fonte: o autor (2023).

No Gráfico 1 é possível observar a quantidade de artigos por ano de publicação:

Gráfico 1 - Quantidade de artigos por ano de publicação.



Fonte: o autor (2023).

Nos quadros a seguir, apresentamos a análise sistemática dos estudos selecionados sobre o conceito de *Data Mesh*. Essa análise permitiu a extração de informações essenciais para compreender e avaliar os diversos aspectos relacionados ao *Data Mesh*, incluindo seus objetivos, métodos utilizados e resultados obtidos.

Por meio desses quadros, buscamos organizar e apresentar de forma clara e concisa as informações extraídas dos estudos selecionados, fornecendo um panorama abrangente sobre o conceito de Data Mesh e suas diversas dimensões.

No Quadro 7, fornecemos uma visão geral das características primárias dos estudos, abordando seus objetivos específicos, os métodos empregados para realizar as pesquisas e os principais resultados alcançados. Essas informações são fundamentais para entender a abrangência e a contribuição individual de cada estudo.

Quadro 7 - Extração de características primárias dos estudos – objetivo, método e resultados.

#	Objetivos	Métodos	Resultados
1	Fornece uma visão holística da governança de dados que é uma síntese dos pontos de vista acadêmicos e profissionais e concluem dando um exemplo de um estudo de caso (<i>Saxo Bank</i>) onde os autores trabalharam em intervenções tecnológicas e culturais para enfrentar os desafios da governança de dados.	Estudo de caso.	De acordo com os autores, uma boa governança de dados federados é crítica porque práticas inadequadas em controles têm impactos negativos significativos nas partes interessadas e na sociedade em geral e, por outro lado, controles excessivos sufocam a produtividade.
2	Os autores propõem a descentralização da arquitetura de dados, onde diferentes funções de negócios (domínios) assumem maior responsabilidade de expor seus dados para uso analítico por toda organização.	Estudo empírico.	A implantação de uma plataforma descentralizada de dados com clusters dentro de cada domínio de negócios trouxe melhor qualidade e interoperabilidade dos dados, orquestrando o gerenciamento de dados e informações em equipes orientadas aos domínios; permitindo o funcionamento dos domínios de forma independente.
3	Os autores apresentaram o mercado de dados corporativos de autoatendimento como uma solução para organizações orientadas a dados para criar e trocar produtos de dados.	Estudo empírico.	Um dos aspectos mais críticos do mercado de dados apresentado incluiu a avaliação automatizada de produtos de dados, que realiza uma estratégia de governança de dados sem intervenção humana.
4	Os autores propõem uma abordagem para transformar um cenário <i>Product Lifecycle Management</i> (PLM) monolítico em um domínio federado e <i>Data Mesh</i> .	Estudo empírico.	Ao mudar, as soluções <i>Product Lifecycle Management</i> (PLM) herdadas monolíticas são substituídas com eficiência por domínios modernos centrados no usuário em um cenário federado. Isso permite que as empresas respondam de forma rápida e flexível aos novos requisitos. Uma camada semântica entre domínios garante a interoperabilidade entre domínios e permite soluções holísticas baseadas em dados e casos de uso analíticos, transformando o cenário fragmentado de <i>Product Lifecycle Management</i> (PLM) em um gráfico de conhecimento vivo. Por fim, uma infraestrutura de <i>Product Lifecycle Management</i> (PLM) independente de domínio aproveita o pensamento de plataforma em todos

			os domínios e garante a interoperabilidade global por meio de padrões globais, bem como a auto soberania do domínio por meio de padrões locais.
5	O estudo analisa o foco emergente repentino da arquitetura <i>Data Mesh</i> como uma nova maneira de pensar sobre os dados. Ele também compara o conceito de malha de dados com <i>data lakes</i> e mostra os princípios associados a essa mudança de abordagem. Por fim, mostra como essa nova estratégia de gerenciamento de dados será uma grande inovação orientada por dados na era atual do <i>Big Data</i> .	Estudo com entrevista estruturada via telefone. Quanti-qualitativo.	O <i>Data Mesh</i> é considerado uma boa estratégia de dados para cientistas e analistas de dados, bem como para organizações, para uma tomada de decisão inteligente e bem-sucedida, aumentando a produtividade e a eficiência para alcançar resultados de negócios rapidamente.
6	O estudo mostra como os dois conceitos são compatíveis em termos de funcionalidade, modelagem de dados, capacidade de evolução e capacidades alinhadas. O <i>data mesh</i> orientado por <i>mask-mediator-wrapper</i> facilita: testes de adoção de baixo risco, prototipagem rápida, padronização e uma garantia de capacidade de evolução.	Estudo hipotético dedutivo.	Apesar do <i>Data Mesh</i> ser um conceito promissor, ainda está em sua infância e, como acontece com qualquer novo conceito, carece de melhores práticas e padrões combinados. Sem dúvida, eles chegarão à medida que a experiência de desenvolvimento do mundo real for reunida pelos primeiros usuários. Os possíveis adotantes do <i>Data Mesh</i> podem vir de várias áreas de negócios e de diferentes tamanhos organizacionais, portanto, seria benéfico para eles executar o <i>Data Mesh</i> de teste antes de se comprometer com a adoção total.
7	O estudo se propõe a resolver o gap do conceito de <i>data mesh</i> – que ainda é novo e carece de insights empíricos do campo. Fornecendo uma compreensão dos fatores motivacionais para introduzir o <i>data mesh</i> , os desafios associados, às melhores práticas, seu impacto nos negócios e possíveis arquétipos.	Entrevista semiestruturada. Quanti-qualitativo.	Entre outras percepções notadas no estudo, a que as organizações têm dificuldades com a transição para a governança federada associada ao conceito de <i>Data Mesh</i> , a mudança de responsabilidade pelo desenvolvimento, fornecimento e manutenção de produtos de dados e a compreensão do conceito geral foi a principal.
8	O autor apresenta o monitoramento inteligente como um produto de dados entre domínios. São fornecidas as principais considerações para projetar um produto de dados de monitoramento inteligente baseado em eventos. O autor apresenta três possíveis domínios necessários para a criação de um sistema de monitoramento inteligente em cada comunidade.	Estudo empírico.	Sistemas de monitoramento inteligentes estão sendo cada vez mais utilizados em muitos países para ajudar a gerenciar eventos de emergência. No entanto, projetar tal sistema é uma tarefa complexa. Tal sistema deve lidar com o volume, variedade e velocidade dos dados modernos.

Fonte: o autor (2023).

No Quadro 8, concentramos nossa atenção nas características dos estudos relacionadas aos benefícios e/ou vantagens do *Data Mesh*, melhores práticas identificadas, casos de uso apresentados, limitações mencionadas e tendências futuras. Essas informações oferecem insights valiosos sobre as possíveis vantagens e desafios da implementação do *Data Mesh*, bem como orientações para práticas eficazes e possíveis direções futuras de pesquisa.

Quadro 8 - Características dos estudos de benefícios e/ou vantagens; melhores práticas; casos de uso; limitações e tendências futuras.

#	Benefícios e/ou vantagens	Melhores práticas	Casos de uso	Limitações	Tendências futuras
1	Dado como benefício a inovação segura e baseada em dados em escala, melhores controles e conformidades e melhores resultados para indivíduos e sociedade.	Boas soluções requerem equipes multidisciplinares diversificadas, ferramentas que as equipes possam usar de forma autônoma, permitindo domínios e modelos de governança escaláveis.	Estudo de caso do banco Saxo.	N/A.	N/A.
2	Foram mencionados benefícios muito similares ao desenvolvimento de software orientado a eventos, onde possui melhor escalabilidade e controle dos dados.	Houve indicações claras de que ter uma equipe de dados multifuncional trabalhando no domínio do negócio junto com as partes interessadas permitiu que as equipes se concentrassem mais claramente nas	Empírico: arquitetura monolítica (<i>Data warehouse</i>) para arquitetura descentralizada.	N/A.	As grandes empresas modernas estão lutando para encontrar uma maneira de escalar porém ainda sem muita confiança, já que a proposta do <i>Data Mesh</i> é extremamente recente e ainda sem muitos padrões bem estabelecidos.

		necessidades, sendo capazes de responder às novas demandas de forma mais ágil e eficaz.			
3	Uma plataforma de dados pode trazer diversos benefícios, como segurança, controle, estabilidade e confiança. Também é mencionado que os produtores dos dados não precisam investir tempo em gerenciar os <i>pipelines</i> para ingestão de dados, já que a plataforma é capaz de deixar isso o mais <i>self-service</i> possível, com escalabilidade e de fácil acesso.	A plataforma segue alguns princípios e boas práticas de desenvolvimento de software, como o uso de tecnologias agnósticas; no core, a plataforma agindo como serviço para toda a companhia, desde analistas até cientistas; multi-tenacidade para os módulos sem custo operacional extra.	Empírico: Desenvolvimento de uma plataforma descentralizada e <i>self-service</i> de dados.	N/A.	Planos futuros de pesquisas e desenvolvimentos no mercado de dados incluem o suporte ao ciclo de vida de um modelo de <i>machine learning</i> , <i>streaming</i> de dados e também a plataforma como ferramenta para engenheiros de dados, fornecendo a transformação e materialização dos dados.
4	A migração da arquitetura de monolítica para descentralizada trouxe muita agilidade nos <i>insights</i> para tomada de decisão, o que é um fator extremamente competitivo por conta do alto dinamismo do mercado dependendo de variáveis internas e externas.	Importância de aplicar abordagens modernas, como uma arquitetura direcionada à domínios.	Empírico: <i>Product Lifecycle Management</i> (PLM) monolítico para domínio federado e <i>Data Mesh</i> .	Possibilidade de acessar dados armazenados em aplicações legado, isso normalmente causa muito tempo e dinheiro da operação.	Com o amadurecimento do <i>Data Mesh</i> , existe a possibilidade de novas filosofias surgirem no cenários dos <i>Product Lifecycle Managements</i> (PLM's), onde a tecnologia se adapta ao usuário, onde os produtos de dados são considerados peças fundamentais no desenvolvimento.
5	Apesar do fator tempo, do dinheiro gasto e dos esforços colocados em <i>data lakes</i> e <i>data warehouses</i> , o <i>Data Mesh</i> foi	Uma boa estratégia para gerenciamento de dados deve ter os 4 pilares: prática (fácil da organização	Entrevista.	O uso da abordagem de entrevista estruturada limitou o número de participantes a apenas 10.	Dados úteis e precisos são absolutamente vitais para que as organizações tomem decisões assertivas, criem planos futuros

	confirmado melhor na área de escala e velocidade. O novo paradigma converte arquiteturas monolíticas, como <i>lakes</i> e <i>warehouses</i> , para um tipo distribuído. Em geral, a malha de dados pode ser explorada para reforçar a qualidade dos dados, melhor governança e fazer uma transformação de alta qualidade com a descentralização da responsabilidade dos dados para diferentes níveis de domínio.	manter), relevante, escalável e integrada.			com base em uma perspectiva analítica e, finalmente, tenham sucesso.
6	A padronização beneficia a capacidade de composição dos sistemas, os sistemas usam as mesmas interfaces para funcionar. Pode-se afirmar que a padronização diminui os graus de liberdade de implementação, mas traz o efeito benéfico de manter os projetos próximos ao conceito original do <i>Data Mesh</i> . Se o <i>Data Mesh</i> for considerado inadequado para a organização, pode ser facilmente desmontado e o sistema analítico reintroduzido.	Apesar de o <i>Data Mesh</i> ser um conceito promissor, ela ainda está em sua infância e, como acontece com qualquer novo conceito, carece de melhores práticas e padrões combinados.	Proposta empírica do uso da arquitetura <i>mask-mediator-wrapper</i> como um driver da <i>Data Mesh</i> .	A arquitetura <i>Mask-Mediator-Wrapper</i> (MMW) não exige que o sistema analítico legado seja desmontado, portanto, uma possível deterioração do serviço é limitada ao período de execução do teste de adoção.	A arquitetura <i>Mask-Mediator-Wrapper</i> (MMW) também pode ser usada para conduzir outras arquiteturas de gerenciamento de dados (por exemplo, data hub, data fabric, data spoke). Pesquisas adicionais também podem explorar a capacidade da arquitetura de conduzir migrações para outras arquiteturas.
7	Os profissionais precisam equilibrar as vantagens com a	Unidade <i>Cross-domain</i> ; capacitar e observar;	Entrevista para melhores práticas.	Para responder à pergunta de pesquisa, foram realizadas	Pesquisas futuras devem investigar os resultados em um

	possível criação de gargalos e silos. Em um ambiente <i>data mesh</i> , o proprietário do produto de dados pode assumir essas responsabilidades adicionais ou compartilhá-las com outro membro da equipe do produto de dados.	vitórias rápidas; Adoção consciente; propriedade dedicada; foco em <i>stewardship</i> .		15 entrevistas semiestruturadas com especialistas. A natureza qualitativa do trabalho resulta em validade quantitativa limitada. No entanto, a abordagem qualitativa foi justificada por se tratar de um novo tema de pesquisa.	nível quantitativo.
8	Latência analítica, segurança, confiabilidade e problemas de dimensionamento levaram a maioria dos projetos de sistemas inteligentes para uma arquitetura multicamada distribuída. Dimensionar cada camada, horizontal e verticalmente, é mais econômico, mais prático e oferece mais flexibilidade no futuro. A arquitetura multicamada ajuda a remover qualquer fonte única de falha no sistema analítico. Embora uma arquitetura multicamada distribuída forneça esses benefícios, ela apresenta um gerenciamento de dados mais complexo do que uma arquitetura centralizada.	N/A.	Proposta empírica: apresenta três possíveis domínios necessários para a criação de um sistema de monitoramento inteligente em cada comunidade.	N/A.	O crescimento de sistemas de monitoramento inteligentes possuem grande complexidade por serem orientados a eventos, então o <i>Data Mesh</i> pode se tornar um grande aliado nesse tipo de situação.

Fonte: o autor (2023).

No Quadro 9, exploramos características relacionadas ao tipo de organização estudada, estratégias adotadas, ferramentas e tecnologias utilizadas, abordagens de governança implementadas e desafios enfrentados. Esses aspectos fornecem uma compreensão mais ampla de como diferentes organizações estão aplicando o conceito de *Data Mesh*, as estratégias empregadas, as ferramentas escolhidas, as abordagens de governança adotadas e os desafios encontrados ao longo do processo

Quadro 9 - Características dos estudos de tipo de organização do estudo; estratégias; ferramentas; tipo de governança e desafios.

#	Tipo de org.	Estratégias	Ferramentas e tecnologias	Abordagem de governança	Desafios
1	Financeira; Banco.	Para inspirar confiança nos dados, a estratégia de governança de dados abordou três aspectos principais: descoberta, segurança e responsabilidade.	Abaixo estão algumas ferramentas e tecnologias utilizadas para aplicar a catalogação e qualidade dos dados: Catálogo de dados: <i>Colibra</i> , <i>Alation</i> , <i>Data.world</i> , <i>Zeenea</i> , <i>Marquez</i> , <i>Amundsen</i> , e <i>Great Expectations</i> ; Qualidade de dados: <i>Informatica IDQ</i> , <i>Trillium DQ</i> , <i>AWS Deequ</i> , <i>Apache Griffin</i> , e <i>LinkedIn Data Hub</i> .	Aplicação da governança federada, onde ela pode ser conquistada aplicando <i>frameworks</i> de catalogação e qualidade de dados em prática.	De acordo com os autores, para alcançar a governança de dados federada, uma organização deve resolver alguns desafios cruciais que são uma mistura de técnicos e sociais, tornando-o tão difícil. Então, colocar negócios e tecnologia na mesma página é muito importante, porque a necessidade de ferramentas de autoatendimento e reorganização é melhor impulsionada pelos negócios, não pela tecnologia. Além disso, a integração perfeita de ferramentas com o ecossistema de dados existente é difícil. As ferramentas disponíveis podem não corresponder exatamente às necessidades organizacionais. Selecionar a ferramenta certa de acordo com as necessidades exclusivas e a integração perfeita torna-se extremamente importante. A cultura e o gerenciamento de mudanças geralmente são a parte mais difícil de qualquer equação. Requer parceria em vários níveis e adesão da alta administração.
2	Manufatura.	A estratégia consiste na migração de uma arquitetura	N/A.	Foi mencionado que após a descentralização e o	A arquitetura centralizada começou a apresentar vários desafios e gargalos para o

		centralizada do tipo <i>data warehouse</i> para uma plataforma descentralizada.		dado ficando na mão dos domínios, eles normalmente conseguem dar mais atenção aos principais pilares da governança, que são o catálogo e a qualidade.	negócio. Com um cenário de aplicativos antigos e monolíticos, os dados de muitos dos sistemas operacionais de origem não eram acessíveis para uso analítico ou de relatórios nem para o <i>data warehouse</i> . Isso ocorreu devido a motivos de segurança, problemas de desempenho ou limitações de rede. Como o <i>data warehouse</i> era, em muitos casos, a única forma de acessar os dados armazenados nos sistemas operacionais, ele era usado tanto para fins operacionais quanto analíticos. Quando vários sistemas operacionais foram integrados ao sistema central, tornaram o modelo de dados do data warehouse muito complexo. Além disso, a lógica de carregamento, bem como a lógica de negócios, eram difíceis de manter.
3	Desenvolvimento de software.	A estratégia é a criação de uma plataforma de dados <i>self-service</i> com o objetivo de atender todos os domínios da organização de forma segura e escalável sem custos operacionais.	Para o desenvolvimento da plataforma uma série de ferramentas foram utilizadas, dentre elas: Ingestão: <i>RESTful de sistemas</i> ; Armazenamento: <i>Amazon S3</i> ; Processamento: <i>Spark, DBT e Databricks</i> ; Orquestração: <i>Airflow, Matillion e Snowflake Tasks</i> ; Visualização: <i>Looker</i> .	Apresenta uma maneira diferente de abordagem à governança federada, removendo a aplicação da governança da interação humana e passando a criar funções automatizadas para a avaliação do produto de dados.	Um dos aspectos mais desafiadores apresentados no mercado de dados é a avaliação automatizada de produtos de dados, que não necessitem de intervenção humana para gerar uma estratégia de governança de dados.

4	Manufatura.	Diferentes estratégias podem ser adotadas para diferentes domínios e capacidades de negócios. Para um domínio <i>Product Lifecycle Management</i> (PLM) pode ser suficiente adquirir um aplicativo de software comercial do mercado, para outro pode ser necessário um desenvolvimento completamente novo. É importante apenas não comprometer os objetivos estratégicos gerais, como interoperabilidade semântica ou usabilidade.	Conceitos gerais foram mencionados, como: Tecnologia web semântica, para melhorar a interoperabilidade dos dados no ciclo de vida do produto; <i>Domain-Driven-Design</i> (DDD): que é o desenvolvimento baseado nos domínios.	O cenário do <i>Product Lifecycle Management</i> (PLM) requer interoperabilidade constante, por isso é necessário uma governança que aborda a descentralização federada, sendo aplicada globalmente nos domínios. Isso inclui sintaxe de <i>schemas</i> , formatação dos dados, modelagem, metadados, segurança, e compliance.	Um dos problemas subjacentes mais críticos nas empresas de manufatura é que a maioria dos sistemas de informação são sistemas legados monolíticos. Os sistemas legados são normalmente de missão crítica, caros de manter devido à falta de documentação adequada e compreensão do sistema, difíceis de integrar com outros sistemas, extremamente difíceis de estender ou modificar, caros e demorados para solucionar problemas e executados frequentemente em hardware obsoleto.
5	Conferência de engenharia e ciência.	Dois entrevistados (20%) afirmaram que nenhuma estratégia está perfeitamente isenta de erros, se não imediatamente, mais tarde no projeto.	N/A.	A governança federada, como um dos pilares do <i>Data Mesh</i> , que é a criação de dados precisos, fatuais e de alta qualidade, por exemplo, uma confiabilidade organizacional.	N/A.
6	Acadêmica.	N/A.	N/A.	Arquitetura <i>Mask-Mediator-Wrapper</i> (MMW).	Construção da plataforma <i>Data Mesh</i> ; Construção do produto de dados com armazenamento específico.
7	Especialistas da indústria.	Alinhamento de conceito e cultura foi a estratégia de melhor impacto na hora da implantação da arquitetura,	N/A.	Os autores reconhecem que a abordagem aos sistemas e arquitetura de gerenciamento de dados de	Mostram limitações nas abordagens de governança federada; Mudança de responsabilidade; Qualidade de metadados;

		segundo entrevistados, pois aumentou significativamente a confiabilidade dos funcionários.		uma organização é altamente individual. Consequentemente, não existe uma abordagem única para todos e as organizações precisam avaliar os trade-offs e aplicar as melhores práticas de acordo com suas necessidades.	Falta de compreensão do conceito do <i>Data Mesh</i> ; Falta de recursos financeiros, técnicos e humanos; Problemas de aceitação e resistência dos líderes.
8	Instituições de saúde.	O domínio ER consome dados como um produto dos domínios de detecção e aviso. Os dados analíticos do domínio de advertência permitem que o domínio ER planeje uma estratégia de saúde mais eficaz para a comunidade.	São utilizadas no caso de uso ferramentas e tecnologias como: <i>Apache Kafka, Apache Sqoop, Amazon S3 e HDFS</i>	É mencionado o conceito de governança federada, porém não menciona sua aplicabilidade no caso mencionado.	N/D.

Fonte: o autor (2023).

5 CONSIDERAÇÕES FINAIS

Com base na análise dos artigos selecionados para esta revisão sistemática sobre *Data Mesh*, é possível destacar alguns pontos importantes.

Primeiramente, observou-se que desde seu surgimento, o conceito de *Data Mesh* vem sendo abordado com uma frequência significativa em diversos artigos. Além disso, foram identificadas aplicações específicas em setores como instituições financeiras, instituições de saúde e indústria de manufaturas, como sistemas de monitoramento inteligentes e gestão do ciclo de vida do produto (*Product Lifecycle Management*).

No que diz respeito aos benefícios do *Data Mesh*, os artigos revisados apresentaram informações claras e explícitas. Destacam-se a melhoria da escalabilidade, governança e agilidade organizacional, decorrentes da quebra de silos e descentralização do gerenciamento de dados.

No entanto, um ponto a ser ressaltado é a lacuna de conhecimento sobre as limitações do *Data Mesh*. Poucos estudos abordaram esse aspecto específico, o que sugere a necessidade de mais pesquisas nessa área para compreender as possíveis desvantagens ou desafios associados à implementação do *Data Mesh*.

Em relação às ferramentas e tecnologias, observou-se que o conceito de *Data Mesh* não exerce um impacto significativo nesses aspectos. As ferramentas e tecnologias mencionadas nos estudos revisados não apresentaram mudanças substanciais, uma vez que o *Data Mesh* é mais uma mudança de paradigma organizacional do que uma transformação tecnológica.

Os desafios identificados estão principalmente relacionados ao entendimento e alinhamento do conceito de *Data Mesh* por parte dos envolvidos. A necessidade de compreensão abrangente do conceito e a aplicação de governança federada foram mencionadas como desafios significativos. Essas questões podem impactar a confiabilidade e a flexibilidade do trabalho realizado, uma vez que a governança em um ambiente descentralizado pode apresentar desafios adicionais.

Em síntese, esta revisão sistemática destaca a emergente discussão em torno do conceito de *Data Mesh*, apontando para sua relevância e benefícios potenciais em áreas específicas de negócios. No entanto, são necessárias mais pesquisas para compreender melhor as limitações e superar os desafios relacionados à implementação e adoção do *Data Mesh*. A compreensão mais aprofundada desses aspectos contribuirá para a evolução do conceito e para orientar as organizações na aplicação eficaz do *Data Mesh* em seus contextos específicos.

REFERÊNCIAS

JOSHI, Divya; PRATIK, Sheetal; RAO, Madhu Podila. **Data Governance in Data Mesh Infrastructures: The Saxo Bank Case Study**, 2021. ICEB 2021 (Nanjing, China). 52. Disponível em: <https://aisel.aisnet.org/iceb2021/52>. Acesso em: 08 jun. 2023.

BODE, Jan et al. **Data Mesh: Best Practices to Avoid the Data Mess**, 2023. Disponível em: <https://arxiv.org/pdf/2302.01713.pdf>. Acesso em: 20 abr. 2023.

DEGHANI, Zhamak. **Data Mesh: A Modern Data Architecture for Competitive Advantage**. ThoughtWorks, 2020. Disponível em: <https://martinfowler.com/articles/data-mesh-principles.html>. Acesso em: 12 mai. 2023.

MORGAN, J. P. **Evolution of Data Mesh Architecture Can Drive Significant Value in Modern Enterprise**, s.d. Disponível em: <https://www.jpmorgan.com/technology/technology-blog/evolution-of-data-mesh-architecture>. Acesso em 15 mai. 2023.

SAMPAIO, Rosana Ferreira; MANCINI, Marisa Cotta. Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. **Brazilian Journal of Physical Therapy**, v. 11, p. 83-89, 2007.

LOUKIALA, Antti et al. Migrating from a centralized data warehouse to a decentralized data platform architecture. In: **Product-Focused Software Process Improvement: 22nd International Conference, PROFES 2021**, Turin, Italy, November 26, 2021, Proceedings 22. Springer International Publishing, 2021. p. 36-48.

MACHADO, Inês; COSTA, Carlos; SANTOS, Maribel Yasmina. Data-driven information systems: the data mesh paradigm shift. 2021. In: **29th International Conference On Information Systems Development**, ISD 2021, Valência, Espanha.

DONCEVI, Juraj; FERTALJ, Kresimir; BRCI, Mario; KOVAC, Mihael. **Mask-Mediator-Wrapper architecture as a Data Mesh driver**. IEEE JOURNAL, VOL. X, NO. Y, SEPTEMBER 2022.

HOOSHMAND, Y.; RESCH, J.; WISCHNEWSKI, P.; PATIL, P. **From a Monolithic PLM Landscape to a Federated Domain and Data Mesh**. International Design Conference – Design 2022. Disponível em: <https://doi.org/10.1017/pds.2022.73>. Acesso em 29 mai. 2023.

JOSHI, Divya; PRATIK, Sheetal; RAO, Madhu Podila. **Data Governance in Data Mesh Infrastructures: The Saxo Bank Case Study**. The 21st International Conference on Electronic Business, Nanjing, China, December 3-7, 2021.

MACHADO, Inês Araújo; COSTA, Carlos; SANTOS, Maribel Yasmina. Data mesh: concepts and principles of a paradigm shift in data architectures. **Procedia Computer Science**, v. 196, p. 263-271, 2022.