

LUCAS DOS SANTOS GOMES

**Descentralização dos ambientes informacionais em sistemas de  
saúde: uma abordagem com data mesh**

São Paulo  
2025

LUCAS DOS SANTOS GOMES

**Descentralização dos ambientes informacionais em sistemas de  
saúde: uma abordagem com data mesh**

**Versão Original**

Monografia apresentada ao PECE – Programa de Educação Continuada em Engenharia da Escola Politécnica da Universidade de São Paulo como parte dos requisitos para a conclusão do curso de MBA em Engenharia de Software.

Área de Concentração: Engenharia de Software

Orientador: Prof. Dr. Alexandre Nardi

São Paulo  
2025

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

### **Catálogo-na-publicação**

Gomes, Lucas dos Santos

Descentralização dos ambientes informacionais em sistemas de saúde: uma abordagem com data mesh / L. S. Gomes -- São Paulo, 2025.

55 p.

Monografia (MBA em Engenharia de Software) - Escola Politécnica da Universidade de São Paulo. PECE – Programa de Educação Continuada em Engenharia.

1.Ciência da Informação 2.Big Data 3.Data Mesh I.Universidade de São Paulo. Escola Politécnica. PECE – Programa de Educação Continuada em Engenharia II.t.

Nome: GOMES, Lucas dos Santos

Título: Descentralização dos ambientes informacionais em sistemas de saúde: uma abordagem com data mesh

Monografia apresentada ao PECE – Programa de Educação Continuada em Engenharia da Escola Politécnica da Universidade de São Paulo como parte dos requisitos para a conclusão do curso de MBA em Engenharia de Software.

Aprovado em:     /     /

Banca Examinadora

Prof(a). Dr(a). \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_

Prof(a). Dr(a). \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_

Prof(a). Dr(a). \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_

## **AGRADECIMENTOS**

Ao PECE – Programa de Educação Continuada em Engenharia, que proporcionou uma experiência acadêmica enriquecedora, possibilitando o aprimoramento profissional e pessoal por meio de um ensino de excelência.

Aos meus pais, que com amor, paciência e dedicação me ensinaram o valor do esforço e da resiliência. Por acreditarem nos meus sonhos, por cada palavra de incentivo e por estarem ao meu lado em cada etapa desta jornada, deixo aqui minha eterna gratidão.

## RESUMO

GOMES, Lucas. **Descentralização dos ambientes informacionais em sistemas de saúde: uma abordagem com data mesh**. 2025. p. 55. Monografia (MBA em Engenharia de Software). Programa de Educação Continuada em Engenharia da Escola Politécnica da Universidade de São Paulo. São Paulo. 2025.

O setor de saúde tem experimentado um crescimento significativo na geração de dados provenientes de diversas fontes, com estimativas indicando que apenas 20% desses dados estão estruturados (GENTNER *et al.*, 2023). O avanço da Internet das Coisas (IoT), impulsionado pela proliferação de dispositivos como *wearables* e sensores inteligentes, ampliou consideravelmente as possibilidades de coleta e análise de dados (SINHA, 2024), fazendo com que, em 2020, o volume de informações geradas no setor ultrapassasse 35 ZB (GENTNER *et al.*, 2023). Nesse contexto, soluções como os data lakes se destacam por sua flexibilidade para lidar com a heterogeneidade dos dados, em comparação aos tradicionais Data Warehouses (GENTNER *et al.*, 2023). Contudo, a centralização dos dados em um único repositório apresenta desafios crescentes de escalabilidade (AWS, 2023), especialmente no setor de saúde, onde a diversidade das análises é vasta, englobando desde o processamento de imagens médicas, como radiografias e tomografias, até a análise genômica (BELLE *et al.*, 2015). Diante dessas limitações da centralização, o conceito de data mesh surge como uma abordagem descentralizada promissora, transferindo o controle e a governança dos dados para especialistas nos domínios de negócios, mais próximos das especificidades de suas áreas (DEHGHANI, 2019). O data mesh propõe a democratização do acesso aos dados, capacitando gestores e tomadores de decisão a utilizar os ativos informacionais de forma mais eficaz, minimizando gargalos operacionais e promovendo maior flexibilidade na infraestrutura. (DEHGHANI, 2023). Este trabalho propõe-se a analisar os princípios fundamentais do data mesh, conforme descritos pela autora Dehghani (2023), adaptando-os ao contexto do setor de saúde. Capacitando gestores e tomadores de decisão a avaliar a viabilidade dessa abordagem em suas organizações, com exemplos práticos e delineando os pré-requisitos necessários para iniciar a descentralização. Conclui-se no final da monografia, que a adoção dessa estratégia, quando alinhada ao plano de expansão da organização, pode mitigar problemas típicos da centralização em sistemas de saúde, como o distanciamento entre as equipes produtoras de informação e a proliferação descontrolada de dados, alcançado pela criação de domínios de dados que atuam simultaneamente como produtores e consumidores. Bem como superando as dificuldades de escalabilidade dos sistemas centralizados com a replicação de padrões e modelos de comunicação que facilita a adição de novos produtos à plataforma de forma estruturada, através da disponibilização de APIs e serviços conectados à rede interligada, tornando a infraestrutura mais flexível. (DEHGHANI, 2023).

Palavras-chave: Data Mesh, Ciência da Informação, Setor de Saúde, Big Data

## ABSTRACT

GOMES, Lucas. **Decentralization of informational environments in healthcare systems: a data mesh approach**. 2025. 55 p. Monografia (MBA em Tecnologia de Software). Programa de Educação Continuada em Engenharia, Escola Politécnica da Universidade de São Paulo, São Paulo, 2025.

The healthcare industry has experienced significant growth in data generation from various sources, with estimates indicating that only 20% of this data is structured (GENTNER et al., 2023). The advancement of the Internet of Things (IoT), driven by the proliferation of devices such as wearables and smart sensors, has considerably expanded the possibilities for data collection and analysis (SINHA, 2024), leading to a data volume exceeding 35 ZB in the sector by 2020 (GENTNER et al., 2023). In this context, solutions like data lakes stand out for their flexibility in handling data heterogeneity compared to traditional Data Warehouses (GENTNER et al., 2023). However, centralizing data in a single repository poses increasing scalability challenges (AWS, 2023), especially in healthcare, where the variety of analyses is vast, encompassing everything from medical imaging processing, such as X-rays and CT scans, to genomic analysis (BELLE et al., 2015). Given these limitations of centralization, the concept of data mesh emerges as a promising decentralized approach, transferring control and governance of data to domain experts who are closer to the specificities of their areas (DEHGHANI, 2019). Data mesh aims to democratize data access, enabling managers and decision-makers to use informational assets more effectively, minimizing operational bottlenecks, and promoting greater flexibility in infrastructure (DEHGHANI, 2023). This study aims to analyze the fundamental principles of data mesh as described by Dehghani (2023), adapting them to the healthcare context. It seeks to empower managers and decision-makers to evaluate the feasibility of this approach in their organizations through practical examples while outlining the prerequisites and planning needed to initiate decentralization. The monograph concludes that adopting this strategy, when aligned with the organization's expansion plan, can mitigate typical problems of centralization in healthcare systems, such as the disconnect between information-producing teams and the uncontrolled proliferation of data. This is achieved by creating data domains that simultaneously act as producers and consumers, and by overcoming the scalability challenges of centralized systems through the replication of patterns and communication models that facilitate the structured addition of new products to the platform. This is further supported by the availability of APIs and services connected to the interconnected network, making the infrastructure more flexible (DEHGHANI, 2023).

Keywords: Data Mesh, Information Science, Healthcare Industry, Big Data

## LISTA DE ILUSTRAÇÕES

	Pág.
Figura 1 – Previsão de dispositivos IoT conectados de 2019 a 2030.....	13
Figura 2 – Princípios do Data Mesh.....	23
Figura 3 – Exemplo de ambiente informacional centralizado em alto nível.....	26
Figura 4 – Decomposição dos domínios de dados em saúde de forma concisa.....	28
Figura 5 – Definição dos produtos de dados do domínio de laboratório.....	32
Figura 6 – Características de uma plataforma de autoatendimento.....	36
Figura 7 – Visão macro da governança federada aplicada ao ambiente.....	44
Figura 8 – Ambiente informacional descentralizado final com Data Mesh.....	46

## LISTA DE TABELAS

Pág.

Tabela 1 – Planejamento para implementar uma plataforma data mesh em saúde	38
Tabela 2 – Comparação dos ambientes centralizado e descentralizado.....	47

## LISTA DE ABREVIATURAS E SIGLAS

AWS	Amazon Web Services
CIOs	Chief information office
DDD	Domain-Driven Design
GCP	Google Cloud Platform
HIPAA	Health Insurance Portability and Accountability Act
IaC	Infraestrutura Como Código
IoT	Internet das Coisas.
LGPD	Lei Geral de Proteção de Dados
SLOs	Service Level Objectives

## SUMÁRIO

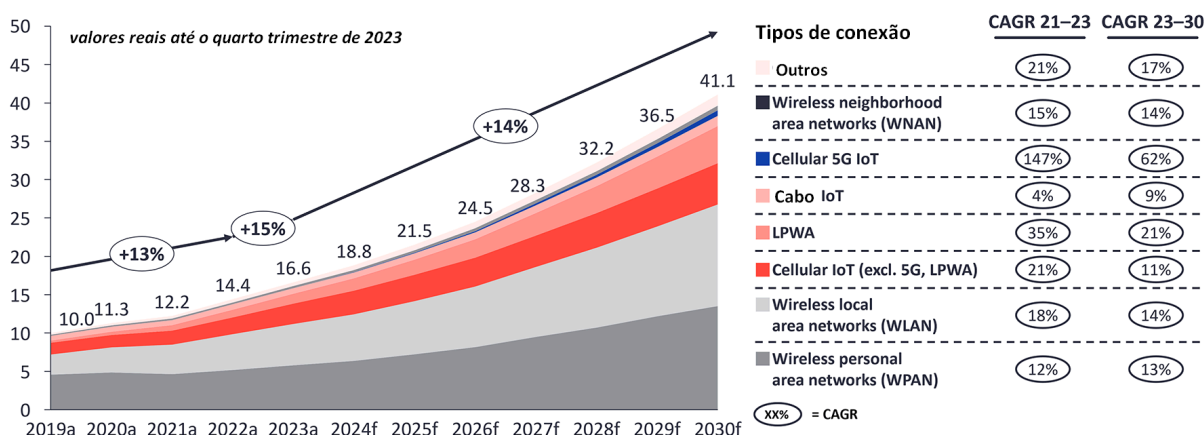
	Pág.
<b>1. INTRODUÇÃO.....</b>	<b>13</b>
1.1 Problema.....	14
1.2 Objetivo.....	16
1.3 Justificativas.....	16
1.4 Método de Pesquisa.....	17
1.5 Estrutura do Trabalho.....	18
<b>2. REVISÃO BIBLIOGRÁFICA.....</b>	<b>19</b>
2.1 Definição do termo big data.....	19
2.1.1 Origem e avanços tecnológicos do big data.....	19
2.1.2 A aplicação do Big Data no setor de saúde.....	20
2.2 Definição de um data lake.....	21
2.2.1 Aplicação e desafios da implementação de data lake no setor de saúde....	21
2.3 Definição do conceito data mesh.....	22
2.3.1 Princípios do data mesh.....	22
2.3.2 Evolução das arquiteturas de dados centralizadas com data mesh.....	24
<b>3. DESCENTRALIZANDO UM AMBIENTE INFORMACIONAL EM SAÚDE COM DATA MESH.....</b>	<b>25</b>
3.1 Ambiente Informacional centralizado: Ponto de Partida.....	25
3.2 Definição dos domínios de negócios e Identificação de seus pré-requisitos...	27
3.2.1 Estruturação e classificação dos domínios de negócio do ambiente.....	28
3.2.2 Aperfeiçoamento contínuo dos domínios de negócios.....	29
3.3 Definição do domínio piloto e seus produtos de dados.....	30
3.3.1 Definição de um produto de dados.....	31
3.3.2 Identificando os produtos de dados do domínio piloto: Laboratório.....	31
3.3.3 Definição da equipe multifuncional e seus papéis no domínio piloto.....	33
3.3.4 Gestão dos produtos e dados centrais da instituição.....	33
3.4 Elaboração do planejamento estratégico de uma plataforma de autoatendimento .....	34

3.4.1	Definição e características de uma plataforma de autoatendimento.....	35
3.4.2	Elaboração da estratégia para implementação de uma plataforma de autoatendimento em saúde.....	38
3.5	Modelo de governança federada.....	41
3.5.1	Definição de governança federada e suas diferenças para a governança de dados tradicional.....	41
3.5.2	Requisitos para aplicar o modelo de Governança em data mesh na saúde	42
3.5.3	Implantação de uma governança federada ao ambiente descentralizado...	43
3.6	Ambiente informacional descentralizado: Conclusão.....	45
<b>4</b>	<b>ANÁLISE DE RESULTADOS .....</b>	<b>47</b>
4.1	Comparação do ambiente centralizado inicial e do ambiente descentralizado gerado.....	47
4.2	Cases de sucesso na implantação do data mesh.....	49
4.2.1	Implementação dos princípios do data mesh na DPG Media.....	50
4.2.2	Implementação dos princípios do data mesh na Globo.....	50
4.3	Considerações do capítulo .....	51
<b>5</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>52</b>
5.1	Contribuições do Trabalho.....	52
5.2	Trabalhos Futuros .....	53
	<b>REFERÊNCIAS .....</b>	<b>54</b>

## 1. INTRODUÇÃO

Nos últimos anos, o setor de saúde tem vivenciado um aumento na quantidade de dados gerados provenientes de uma diversidade crescente de fontes. Estima-se que aproximadamente 20% desses dados estejam em formato estruturado (GENTNER *et al.*, 2023). A popularização de dispositivos *wearables*, que monitoram continuamente os indivíduos ao longo do dia, assim como sensores inteligentes presentes em empresas, hospitais e residências, fazem parte de um ecossistema conhecido como internet das coisas (IoT), o que ampliou significativamente as possibilidades de coleta e análise de dados, conforme ilustrado na figura 1. De acordo com especialistas, esse fenômeno resultou em uma produção de dados no setor de saúde que, em 2020, ultrapassou 35 ZB.(GENTNER *et al.*, 2023).

Figura 1 – Previsão de dispositivos IoT conectados de 2019 a 2030



Fonte: Sinha (2024)

Diante desse cenário, a aplicação de estratégias de armazenamento de dados e tecnologias para o processamento de informações baseadas nos princípios do big data torna-se justificável. O exemplo clássico dos três Vs (Variedade, Volume e Velocidade) é facilmente identificado nesse contexto. A variedade refere-se à diversidade de origens e formatos de dados, incluindo dados estruturados, semiestruturados e não estruturados; o volume diz respeito à imensa quantidade de dados gerados e a velocidade refere-se à rapidez com que os dados precisam ser processados. (ORACLE, 2024).

“Big data são ativos de informação de alto volume, alta velocidade ou alta variedade que exigem formas inovadoras e econômicas de processamento de informações que permitam melhor percepção, tomada de decisões e automação de processos”. (GARTNER, 2024, p. 1, tradução nossa).

Em resposta a essas características, soluções como data lakes têm se mostrado eficazes, proporcionando maior flexibilidade ao lidar com a variedade de dados, em comparação com data warehouses tradicionais, bem como, serem soluções mais econômicas e escaláveis para o armazenamento de grandes volumes de dados. Permitindo que as infraestruturas acompanhem a evolução das análises necessárias. (GENTNER *et al.*, 2023).

A adoção dessas tecnologias para o armazenamento de informações em um único ponto de controle, pode influenciar empresas do setor de saúde a adotar um paradigma centralizado na gestão e processamento de seus dados. Nesse modelo, todas as informações relevantes são armazenadas em um único repositório, assim como seu gerenciamento, o que facilita o controle de acesso e a análise. No entanto, considerando a diversidade de análises nesse setor específico de saúde, que incluem desde o processamento de imagens médicas, como radiografias e tomografias, até a análise genômica, (envolvendo o estudo de aproximadamente 30.000 a 35.000 genes por indivíduo). (BELLE *et al.*, 2015). Faz com que, a abordagem centralizada comece a apresentar alguns problemas e pontos de estresse, principalmente nesse cenário, com domínios de dados densos, com uma variedade de origens e formatos distintos e múltiplas partes interessadas como clínicas, pacientes, hospitais, entre outros. (DEHGHANI, 2019).

## **1.1 Problema**

À medida que os domínios de dados ficam mais densos, com volume e riqueza maior de informações, as organizações começam enfrentar problemas em escalar esse grande repositório centralizado (AWS, 2023). Alguns pontos de pressão começam a surgir nessa arquitetura, como exemplo podemos citar:

A proliferação dos dados dentro das instituições torna cada vez mais difícil padronizá-los e armazená-los em um único local. À medida que os dados se espalham, a capacidade de centralizá-los em um único ponto central, torna-se limitada. Isso reduz a velocidade e a flexibilidade para integrar novas fontes de informação nesse ecossistema dinâmico. (DEHGHANI, 2019).

Com uma equipe central composta por diversos profissionais técnicos separados por tecnologias específicas, que são responsáveis por gerenciar esse ambiente unificado, surgem novos desafios. Um dos principais problemas é o distanciamento desses especialistas em relação à natureza específica do domínio que estão atendendo. Sem uma compreensão profunda e motivação clara quanto às necessidades reais do negócio, suas ações podem parecer desconectadas. (DEHGHANI, 2019). Além disso, concentrar toda a responsabilidade de edição em uma única equipe diminui a agilidade da organização. Cada nova necessidade de análise exige a reformulação de pipelines inteiras, que transformam e padronizam os dados. Isso sobrecarrega a equipe, que não apenas precisa absorver novas informações, mas também gerenciar todo o ecossistema, criando gargalos operacionais. (AWS, 2023).

Por se tratar de dados sensíveis e pessoais, uma barreira encontrada no paradigma centralizado aplicado em saúde está relacionada à legislação. Em diversos países, existem leis e regulamentações que determinam como os dados devem ser armazenados e processados, como a Lei Geral de Proteção de Dados (LGPD) no Brasil e a *Health Insurance Portability and Accountability Act* (HIPAA) nos Estados Unidos, exigindo muitas vezes que os dados sejam processados e tratados localmente, dentro da jurisdição em que foram coletados, dificultando a unificação de dados globais ou com regiões com diferentes regulamentações, bem como, exigindo a implementação de políticas e normas específicas, nos quais, a não conformidade gera multas e penalizações para organização. (BRASIL, 2018). (ESTADOS UNIDOS, 1996).

## 1.2 Objetivo

Diante dos cenários descritos, o conceito de data mesh, com sua proposta descentralizadora, surge como uma alternativa promissora. Esse modelo busca resolver os problemas inerentes à centralização da informação, propondo uma abordagem onde os domínios de negócios assumem a responsabilidade pela geração, gestão e governança de seus próprios dados, não sendo uma proposta totalmente contrária à centralização, mas sim uma mudança na forma como os dados são organizados e gerenciados pela instituição.

Abordando neste trabalho principalmente a democratização do ativo de dados na organização, capacitando gestores e tomadores de decisão sobre os princípios que regem essa nova abordagem e os guiando estrategicamente para uma avaliação de viabilidade em sua organização. Apresentando formas de transferir o controle e gerenciamento desse dado para os especialistas dos domínios, que estão próximos do negócio, bem como, aumentando a flexibilidade da infraestrutura, permitindo que os domínios repliquem técnicas do sistema central. Por fim, buscando formas de aumentar a segurança e conformidade, implementando políticas dentro dos domínios e entre os mesmos. (AWS, 2023).

Resumindo seu objetivo, conforme a definição do data mesh apresentado pela a autora Dehghani (2023, p. 43). “uma abordagem sociotécnica descentralizada para compartilhar, acessar e gerenciar dados analíticos em ambientes complexos e de grande escala - dentro ou entre organizações.”

## 1.3 Justificativa

O conceito de data mesh tem ganhado crescente atenção em discussões acadêmicas e empresariais devido à sua proposta inovadora de descentralizar o gerenciamento de dados. Esse crescimento pode ser atribuído a tendências como a utilização de dados como produto e plataformas de autoatendimento para criar e consumir dados dentro dos domínios. Grandes provedores de tecnologia, como a *Amazon Web Services (AWS)* e *Google Cloud Platform (GCP)*, têm desenvolvido

soluções alinhadas ao modelo de data mesh para atender à crescente demanda de empresas que buscam democratizar o acesso e a governança de seus dados, como visto nos artigos do Shankar *et al.* (2021) e Google (2024a).

Essa tendência tem impulsionado discussões sobre novos paradigmas no gerenciamento do ativo de dados nas instituições. Atualmente, o conceito de data mesh está disseminado, com diversos artigos e livros publicados, disponibilizados para venda em plataformas digitais, entre essas obras, podemos citar o livro da autora Dehghani (2023), que é o 22º mais vendido globalmente na categoria de base de dados. (AMAZON, 2024)

Dessa forma, esse trabalho tem como intuito ou objetivo específico de apresentar uma análise e planejamento estratégico desse novo paradigma sociotécnico a gestores e tomadores de decisão, demonstrando como ele pode mitigar os desafios operacionais, regulamentares e de negócios de empresas centralizadas na área da saúde e orientar estrategicamente para atingir a descentralização.

#### **1.4 Método de Pesquisa**

A pesquisa foi desenvolvida a partir de uma abordagem qualitativa, que combina a revisão bibliográfica com a análise de experiência prática. A revisão bibliográfica foi realizada por meio da leitura e análise de artigos acadêmicos, livros e outros materiais relevantes sobre o conceito de data mesh e sua aplicação em ambientes descentralizados em saúde. Foram priorizadas fontes atualizadas, incluindo publicações de autores renomados, como Zhamak Dehghani, uma das principais especialistas no tema, além de materiais produzidos por grandes empresas de tecnologia, como Amazon, Google e Oracle, que discutem os princípios e benefícios dessa abordagem para a gestão de dados.

Além disso, a pesquisa também se baseou em experiência prática do autor, que aplicou os conceitos de data mesh em contextos específicos, com foco em instituições de saúde. A análise incluiu o estudo de casos reais de implementação em hospitais, a fim de ilustrar a viabilidade dos princípios do data mesh nesses cenários, destacando tanto os desafios enfrentados quanto os resultados obtidos.

## 1.5 Estrutura do Trabalho

O Capítulo 1 INTRODUÇÃO apresenta as motivações, o objetivo, as justificativas, método de pesquisa e a estrutura do trabalho.

O Capítulo 2 REVISÃO BIBLIOGRÁFICA apresenta uma revisão dos estudos mais relevantes sobre o tema de data mesh, Big Data e data lake, destacando sua evolução, fundamentos teóricos e aplicações práticas, com ênfase no setor da saúde.

O Capítulo 3 DESCENTRALIZANDO UM AMBIENTE INFORMACIONAL EM SAÚDE COM DATA MESH explora os princípios do data mesh e como adaptá-los ao setor de saúde. O capítulo analisa um cenário de gestão centralizada de dados e propõe um planejamento estratégico para avaliar a viabilidade de sua descentralização, com o objetivo de capacitar gestores a implementar essa abordagem em suas instituições.

O Capítulo 4 ANÁLISE DE RESULTADOS descreve os resultados obtidos ao aplicarmos os princípios data mesh no ambiente centralizado inicial, comparando os aspectos em ambos os cenários e avaliando se os problemas apresentados se mantêm, além de apresentar cases da implementação com sucesso no mercado desse novo paradigma, com os resultados obtidos e os desafios encontrados.

O Capítulo 5 CONSIDERAÇÕES FINAIS descreve as conclusões gerais sobre os conceitos defendidos na monografia, bem como, as contribuições e oportunidades de trabalhos futuros sobre o tema.

REFERÊNCIAS relacione as fontes consultadas durante a elaboração da monografia, garantindo a transparência e credibilidade da pesquisa.

## **2. REVISÃO BIBLIOGRÁFICA**

Este capítulo tem como objetivo apresentar uma revisão dos estudos mais relevantes sobre os temas de data mesh, big data e data lake, destacando sua evolução, fundamentos teóricos e aplicações práticas, com ênfase no setor da saúde. A revisão abordará as principais publicações acadêmicas, livros, artigos técnicos e estudos de caso que sustentam o desenvolvimento das abordagens apresentadas. Para garantir uma análise clara, o conteúdo está organizado em subseções que permitam explorar cada conceito de forma completa, fornecendo um embasamento teórico sólido para as propostas apresentadas nesta monografia.

### **2.1 Definição do termo big data**

O conceito de Big Data refere-se a formas inovadoras e econômicas para processamento e armazenamento de informações, categorizado pelos '3Vs' o volume de dados, a variedade de formatos e a necessidade de velocidade para o processamento dessas informações (GARTNER, 2024). A combinação desses três aspectos exige a criação de novos modelos e tecnologias, os quais são fundamentais para o desenvolvimento e aplicação do conceito de Big Data.

#### **2.1.1 Origem e avanços tecnológicos do big data**

Embora esse conceito seja recente, suas origens remontam às décadas de 1960 e 1970, com o surgimento dos primeiros centros de dados e bancos relacionais, que marcaram o início da organização sistemática de dados em grande escala. Esses avanços iniciais estabeleceram as bases para o desenvolvimento de soluções mais escaláveis nas décadas seguintes. O avanço de tecnologias como o Hadoop, que se tornou um marco na capacidade de processamento distribuído, e o crescimento das soluções NoSQL, a partir de 2005, transformaram profundamente a forma como grandes volumes de dados podiam ser armazenados e analisados, permitindo novos níveis de eficiência. (ORACLE, 2024).

Com o surgimento da Internet das Coisas (IoT), a interconexão de dispositivos passou a gerar uma quantidade de dados sem precedentes, e o desenvolvimento contínuo de técnicas de aprendizado de máquina trouxe novas possibilidades analíticas, enriquecendo ainda mais o ecossistema de big Data. Essas inovações expandiram consideravelmente as possibilidades de aplicação do Big Data em diferentes setores, desde a indústria até a saúde (ORACLE, 2024), tendo sua popularização na computação em nuvem, pois, tornou-se possível escalar essas soluções de forma elástica, atendendo à demanda crescente. Além disso, os bancos de dados gráficos passaram a desempenhar um papel crucial, oferecendo análises eficazes. Essas inovações tecnológicas, especialmente associadas à nuvem, têm impulsionado a transformação de diversos setores, tornando as soluções de big data cada vez mais acessíveis e eficientes (ORACLE, 2024).

### **2.1.2 A aplicação do Big Data no setor de saúde**

O uso de Big Data na saúde é cada vez mais reconhecido como essencial para melhorar os resultados clínicos e a eficiência dos sistemas de saúde. De acordo com Belle *et al.* (2015), a análise de grandes volumes de dados pode superar falhas nos sistemas atuais, como a dificuldade em integrar e utilizar dados de diferentes fontes; embora haja grandes investimentos em saúde, a falta de uma abordagem integrada prejudica os resultados. O Big Data surge como uma solução para melhorar a qualidade dos cuidados oferecidos e a tomada de decisão.

Além disso, o estudo de Belle *et al.* (2015) sugere que ao integrar dados clínicos, genômicos e de imagens médicas, o Big Data pode melhorar a precisão dos diagnósticos e reduzir o tempo necessário para conclusões. O uso de técnicas como aprendizado de máquina também permite uma análise mais rápida e precisa, podendo reduzir custos, melhorar a relação entre pagamento e desempenho e ampliar o acesso ao atendimento médico, além de contribuir para a personalização do tratamento a longo prazo.

## 2.2 Definição de um data lake

Os data lakes, ou Lagos de Dados, são sistemas de armazenamento que permitem a coleta e análise de grandes volumes de dados brutos, estruturados e não estruturados, com origem de diversas fontes. Diferentemente dos Data Warehouses, que organizam os dados de forma estruturada e consolidada, os data lakes mantêm a flexibilidade de armazenar dados em sua forma original, sem a necessidade de um pré-processamento. Essa abordagem permite que as organizações capturem dados em tempo real e executem análises complexas de acordo com suas necessidades. (GENTNER *et al.*, 2023).

Uma das principais características dos data lakes é sua escalabilidade, permitindo que os dados sejam armazenados de maneira eficiente em grande volume, sem as limitações dos modelos de bancos de dados tradicionais. Além disso, os data lakes são flexíveis, pois podem acomodar diferentes tipos de dados, como logs de sistemas, vídeos, imagens, dados de sensores, redes sociais, entre outros (GENTNER *et al.*, 2023). Com isso, as empresas conseguem explorar dados em formatos variados e aproveitar oportunidades de análises mais amplas, como a aplicação de algoritmos de aprendizado de máquina e inteligência artificial para gerar *insights*. (BELLE *et al.*, 2015)

### 2.2.1 Aplicação e desafios da implementação de data lake no setor de saúde

Na saúde, os data lakes são usados para centralizar e integrar dados com origem de diversas fontes, criando uma visão holística do paciente. Isso permite melhorias na coordenação de cuidados, reduzindo erros médicos e permitindo tratamentos personalizados baseados em uma análise mais precisa (GENTNER *et al.*, 2023). Além disso, os data lakes desempenham um papel crucial em análises preditivas, onde os dados de diferentes fontes, como registros clínicos, dados genômicos e de dispositivos de monitoramento, podem ser usados para prever resultados e otimizar intervenções. (BELLE *et al.*, 2015). Na pesquisa médica, eles possibilitam a exploração de grandes volumes de dados não estruturados, acelerando a descoberta de novos tratamentos e inovações terapêuticas. (GENTNER *et al.*, 2023).

No entanto, o uso de data lakes na saúde enfrenta vários desafios. Conforme apresentado no estudo Gentner *et al.* (2023), destaca-se a importância da governança de dados, já que a natureza desestruturada dos dados pode afetar sua qualidade e integridade, comprometendo a confiabilidade das análises realizadas. Além disso, questões relacionadas à segurança e privacidade são apresentadas no estudo com grande preocupação, dada a sensibilidade dos dados de saúde, que exigem medidas de proteção e conformidade com regulamentações como a HIPAA nos Estados Unidos e a LGPD no Brasil. Outro desafio significativo mostrado, é a integração de dados provenientes de diferentes fontes e formatos, o que exige tecnologias e processos que garantam a consistência e a acessibilidade dos dados para análise.

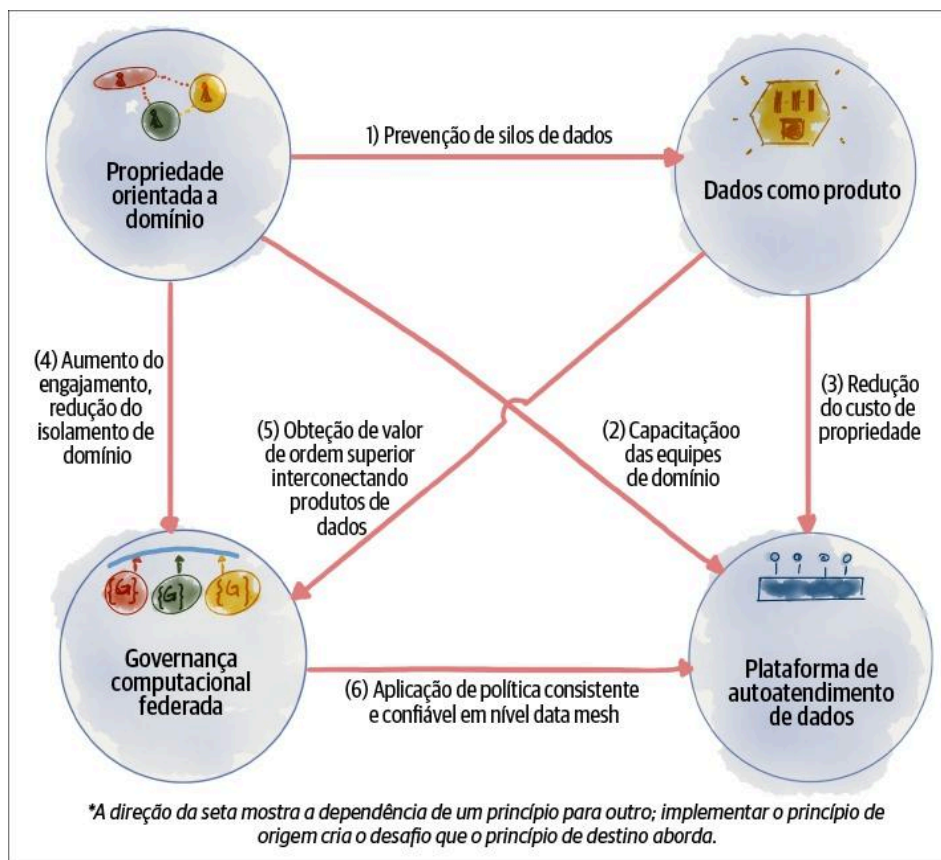
### **2.3 Definição do conceito data mesh**

O conceito de data mesh foi introduzido por Zhamak Dehghani em 2019, como uma solução para os desafios que surgem ao tentar escalar a arquitetura tradicional de dados centralizados e aprofundado através do livro *Data Mesh: Delivering Data-Driven Value at Scale* lançado em 2020. Tradicionalmente, organizações dependem de um Data Warehouse ou data lake centralizado, onde dados de todas as fontes são reunidos e processados de forma unificada. Com o aumento da complexidade dos dados e da quantidade de fontes distribuídas, os modelos centralizados começam a enfrentar limitações de escalabilidade, flexibilidade e capacidade de resposta às necessidades dinâmicas dos negócios. (DEHGANI, 2019). O data mesh surgiu como um paradigma sociotécnico, oferecendo uma abordagem para tratar dados como um produto e aplicar práticas de governança mais distribuídas, federando a responsabilidade de criação e gestão aos domínios de negócios.

#### **2.3.1 Princípios do data mesh**

O data mesh é fundamentado em quatro pilares principais, conforme apresentado pela autora Dehghani (2023), que tem como objetivo transformar a maneira como as organizações estruturam e gerenciam seus dados em larga escala, a figura 2 apresenta os princípios e sua explicação.

Figura 2 – Princípios do data mesh



Fonte: Dehghani (2023, p. 50)

O primeiro pilar é a descentralização dos domínios de dados. Em vez de concentrar todos os dados em um único repositório, o data mesh organiza os dados em domínios correspondentes às unidades de negócios ou funções específicas dentro da organização. Cada domínio é responsável por gerenciar e disponibilizar seus próprios dados de forma independente, oferecendo maior flexibilidade e controle local sobre a qualidade e a governança. (DEHGhani, 2023)

O segundo pilar trata dos dados como um produto. Isso implica que os dados não devem ser vistos como um subproduto de sistemas, mas como um produto completo, com usuários finais que os consomem da mesma forma que outros produtos na organização. Cada domínio deve criar, documentar e manter seus dados com qualidade e acessibilidade, garantindo que os consumidores possam confiar na consistência dos dados oferecidos. (DEHGhani, 2019)

O terceiro pilar é a criação de uma plataforma de dados auto-serviço. O data mesh propõe o desenvolvimento de uma plataforma que permita aos domínios desenvolverem, acessarem e consumirem dados de maneira independente. Esta plataforma deve ser escalável, facilitando o processamento e a análise dos dados, além de possibilitar que os domínios compartilhem dados sem a necessidade de uma equipe centralizada de TI. O objetivo é garantir que qualquer equipe de dados tenha autonomia para operar sem barreiras. (DEHGHANI, 2023)

Por fim, o quarto pilar é a governança federada. Embora os domínios tenham autonomia, é necessário que sigam padrões comuns de qualidade, segurança e conformidade. A governança federada garante que, mesmo com a gestão descentralizada, os dados possam ser usados de forma consistente e segura em toda a organização. (DEHGHANI, 2023)

### **2.3.2 Evolução das arquiteturas de dados centralizadas com data mesh**

Conforme Dehghani (2019) defende, os data lakes representam a segunda geração das arquiteturas de dados, centralizando grandes volumes de dados brutos, mas com desafios de governança, qualidade e escalabilidade. O data mesh, por sua vez, surge como uma evolução ao propor uma abordagem descentralizada para a gestão de dados, distribuindo a responsabilidade para diferentes domínios de negócios. Cada domínio é responsável por seus próprios dados, mas dentro de um modelo de governança federada, garantindo sua consistência. Essa abordagem descentralizada se inspira no conceito de Domain-Driven Design (DDD), que também distribui responsabilidades por domínios de negócios, permitindo maior autonomia e especialização nas equipes. (DEHGHANI, 2023). O data mesh é considerado uma solução mais escalável e ágil para grandes organizações ou aquelas com dados complexos, superando as limitações do data lake. No entanto, ele reconhece a importância da centralização em algumas áreas e muda os paradigmas sobre como trabalhar com arquiteturas centralizadas, oferecendo uma nova perspectiva para a gestão e governança de dados em larga escala.

### **3. DESCENTRALIZANDO UM AMBIENTE INFORMACIONAL EM SAÚDE COM DATA MESH**

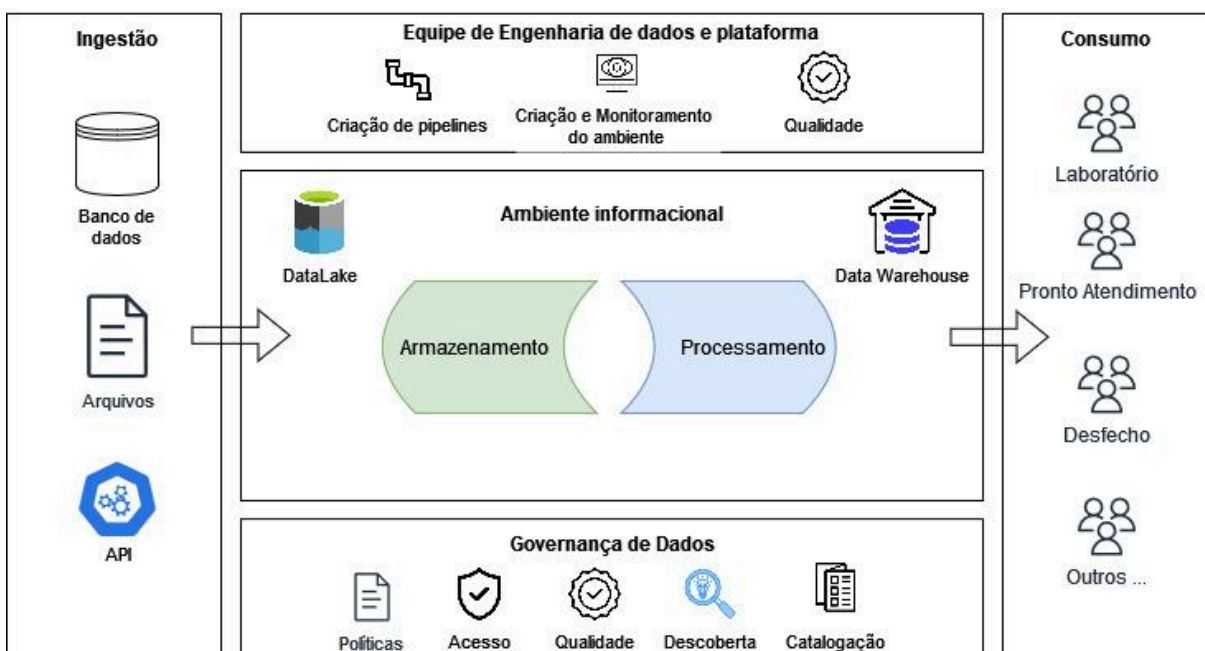
Essa seção tem como objetivo analisar os princípios fundamentais da abordagem data mesh a partir dos conceitos descritos pela autora Dehghani (2023), sendo eles, propriedade de domínio, dados como produtos, plataforma de autoatendimento de dados e governança federada. Buscando efetuar uma adaptação desses princípios ao contexto específico do setor de saúde. Para isso, será analisado um cenário que destaca os desafios identificados nessa abordagem e apresenta estratégias para superar essas barreiras. Além disso, serão discutidas as mudanças necessárias nas estruturas hierárquicas da organização, para garantir uma governança eficiente dos produtos de dados mapeados.

O ponto de partida será uma arquitetura de dados centralizada tradicional, a partir da qual, serão propostas as transformações necessárias para promover a descentralização dos dados. O foco será no aumento da responsabilidade dos domínios de dados, viabilizando a expansão de organizações em saúde sem os problemas típicos enfrentados no modelo atual, objetivando potencializar o uso de dados na instituição e análises desenvolvidas.

#### **3.1 Ambiente Informacional centralizado: Ponto de partida**

Na figura 3, apresentamos uma arquitetura centralizada de dados inspirada em trabalhos reais na área da saúde, dessa forma podemos identificar os principais pontos de pressão na centralização dos dados para esse cenário específico, caracterizado por domínios de dados densos, dinâmicos, com alta demanda de expansão e potencializado com a presença de múltiplas partes interessadas que estão constantemente solicitando novas ingestões e análises de dados.

Figura 3 – Exemplo de ambiente informacional centralizado em alto nível



Fonte: De autoria própria

A figura acima apresenta uma arquitetura no qual centraliza o processo de transformação dos dados, separando as atuações entre equipes especializadas e suas tecnologias, processando e armazenando os dados em um ambiente unificado, assim como, o seu gerenciamento. Podemos verificar claramente os papéis e responsabilidades de cada equipe e reparar que os domínios de negócios apenas consomem as informações disponibilizadas pelos times de engenharia e governança de dados.

A equipe de governança de dados detém todo o poder e responsabilidade para controlar os acessos, garantir a qualidade desse ativo, definir as políticas e normas desse ambiente e etc, assim como, a equipe de engenharia de dados se responsabiliza pelo monitoramento, criação de pipelines para ingestão de novas informações nesse ambiente entre outros papéis.

Dessa forma, alguns obstáculos começam a ser apresentados, como exemplo podemos citar o distanciamento dessa equipe especializada em relação ao domínio que estão atendendo e a sobrecarga de responsabilidades em um único time, conforme já listados na seção de Introdução deste trabalho. Logo a aplicação de

uma arquitetura data mesh, se torna válida quando alinhada com o objetivo de expansão da organização e difusão de uma cultura voltada a dados.

### **3.2 Definição dos domínios de negócios e Identificação de seus pré-requisitos**

Para iniciarmos a decomposição do ambiente informacional apresentado, é fundamental definir os domínios de negócios e suas respectivas responsabilidades. Diferentemente de um ambiente unificado, onde a delimitação ocorre com base na tecnologia utilizada e em equipes especializadas, essa definição se orienta pelo objetivo e pelo resultado almejado (DEHGHANI, 2023). Assim, um domínio de negócio pode ser compreendido como uma unidade voltada para um objetivo específico, otimizada para alcançar esse resultado, podendo ser classificados em três categorias: alinhados à fonte, agregados de domínio e alinhados ao consumidor (DEHGHANI, 2023). Essa classificação facilita a identificação das fronteiras de cada domínio, evitando duplicidades de trabalhos e possíveis disputas de interesses sobre os produtos de dados.

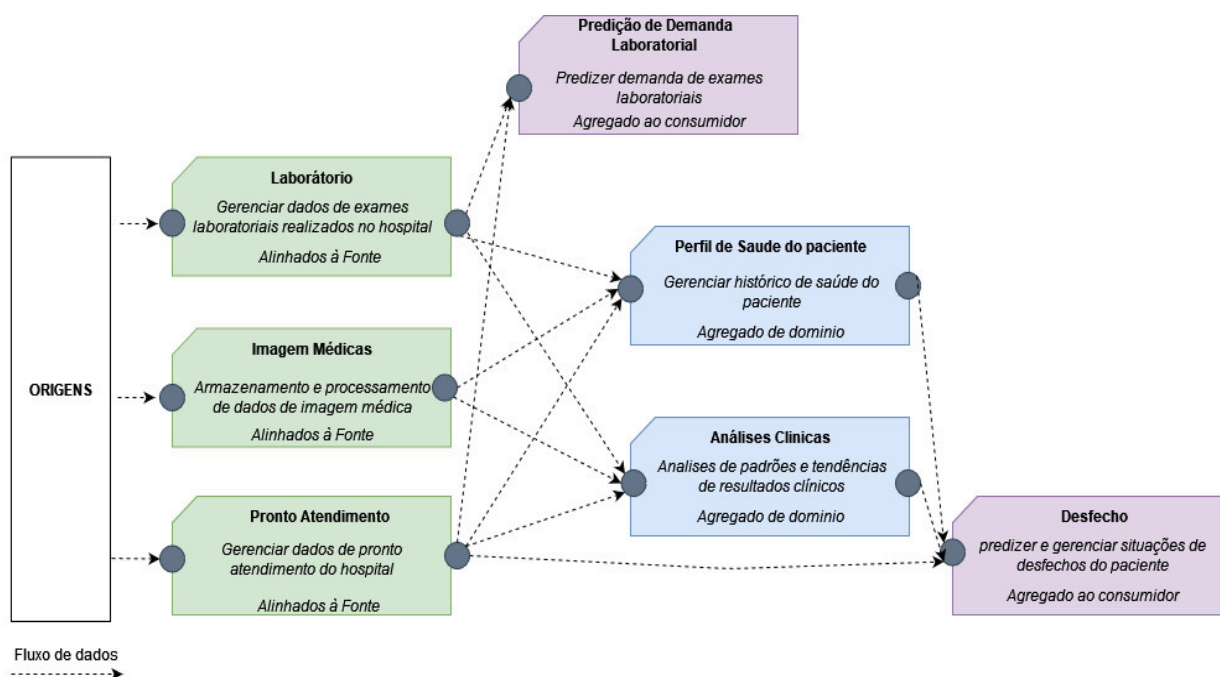
Portanto, ao mapearmos os objetivos de negócios para identificar os domínios da instituição é necessário levarmos em consideração a estrutura das unidades organizacionais da empresa, tendo a liberdade de adaptar o conceito a esse modelo, bem como, respeitar as definições de negócios definidas pelas áreas, entendendo como é o fluxo dos dados dentro dessa estrutura e as correlações de análises e produtos que cada área reporta.

Dessa forma, um pré-requisito para elaboração de uma estrutura descentralizada é a disseminação de uma cultura orientada a dados e domínio, sendo necessário o entendimento claro da responsabilidade e poderes que cada área de negócio possui sobre o produto de dados que gerencia. Assim, workshops, reuniões entre outras táticas devem ser utilizadas para incentivar um ambiente colaborativo e se necessário mudanças estruturais devem ocorrer para que a empresa possa adotar esse novo paradigma sociotécnico.

### 3.2.1 Estruturação e classificação dos domínios de negócio do ambiente

A estruturação e a classificação dos domínios é uma demanda dinâmica, dado que novos produtos são frequentemente adicionados, ajustados ou descontinuados. Nesse contexto, a definição inicial deve priorizar as unidades organizacionais mais consolidadas e alinhadas à estratégia organizacional da instituição e ser melhorada posteriormente. Na figura 4, é apresentada uma proposta concisa de definição e classificação dos domínios de negócios em saúde, decompondo o ambiente centralizado apresentado anteriormente com unidades organizacionais reais de um hospital. Essa abordagem visa materializar os conceitos discutidos, facilitando a compreensão do que constitui um domínio de negócio e como ele se relaciona com as operações hospitalares.

Figura 4 – Decomposição dos domínios de dados em saúde de forma concisa



fonte: De autoria própria

Observa-se claramente que cada domínio de negócio possui objetivos bem definidos. Além de ter o interesse em garantir a qualidade e a atualização dos dados gerados, cada domínio assume a responsabilidade de compartilhá-los de maneira eficaz com a instituição. Essa abordagem promove maior autonomia, transformando

os domínios de negócio de meros consumidores de informações em produtores ativos, que contribuem diretamente para o ambiente informacional da organização.

O domínio de negócio de Laboratório (alinhado à fonte) é responsável por gerar informações a partir de sistemas transacionais e de equipamentos médicos laboratoriais, além de manter um armazenamento histórico e disponibilizado esses dados em tempo real. Essa estrutura permite que o domínio de Perfil de Saúde do paciente (agregado de domínio) possa focar no consumo de informações confiáveis, organizando seus ETLs apenas nas agregações necessárias de acordo com suas necessidades de análises e não na extração da origem. Com essa base sólida, o domínio de Desfecho (alinhado ao consumidor) consegue rodar modelos de inteligência artificial e aprendizado de máquina voltado ao *stakeholder* médico, apresentando as possibilidades de desfecho daquele paciente, utilizando dados de alta qualidade, sem a necessidade de construir todo o pipeline de tratamento dos dados desde a origem.

Cada domínio possui fronteiras bem definidas, o que permite agilidade nas operações. Quando surge a necessidade de realizar uma nova análise com dados de um equipamento adicional, o domínio de Perfil de saúde do paciente solicita ao domínio de Laboratório a inclusão desse novo produto. Em seguida, essa análise é disponibilizada para a equipe de Desfecho, que é responsável por ajustar seus modelos conforme necessário. Dessa forma, é evitada a necessidade de edições extensivas em pipelines completas por um único time especializado que está distante do negócio, o que poderia criar gargalos no processo, além de ser possível adicionar facilmente novos domínios a esse ecossistema.

### **3.2.2 Aperfeiçoamento contínuo dos domínios de negócios**

A definição dos tratamentos dos domínios de negócios é uma demanda cíclica, pois o ambiente dinâmico exige a criação de novos domínios e a adaptação de outros. Quando um domínio se torna excessivamente complexo com múltiplos produtos de dados, é realizada uma avaliação para considerar seu replanejamento. Para que esse processo seja eficaz, é fundamental que a empresa tenha uma cultura

orientada a dados e que cada domínio esteja alinhado à estratégia escolhida pela instituição, o que requer a participação ativa de todos os envolvidos.

### **3.3 Definição do domínio piloto e seus produtos de dados**

Após o mapeamento inicial dos domínios de negócio e sua redefinição estratégica voltada para uma cultura orientada a dados, uma abordagem para iniciar o processo de planejamento estratégico do data mesh na instituição é a partir da definição de um domínio piloto a ser desenvolvido. A escolha desse domínio será baseada em critérios de prioridade, considerando sua relevância ou alinhamento com a estratégia organizacional. O domínio piloto será o primeiro a ser integrado à plataforma de autoatendimento e utilizar uma nova metodologia de governança federada, desenvolvidas posteriormente neste trabalho, possibilitando a construção de uma nova arquitetura de dados paralelamente ao modelo centralizado atualmente em operação.

Essa definição possibilita que a instituição realize a migração para uma arquitetura distribuída de forma gradual, por meio de ciclos de implementação que incluem aprendizado contínuo. Garantindo que novos domínios sejam incorporados de maneira progressiva e com uma complexidade menor, enquanto a organização se mantém competitiva no mercado.

Para fins de exemplificação e melhor entendimento dos conceitos abordados neste tópico sobre os produtos de dados, o domínio de partida escolhido foi o domínio de laboratório. A escolha deve-se à sua relevância estratégica, complexidade e particularidade, caracterizada por uma elevada interoperabilidade de sistemas com múltiplas partes interessadas no seu consumo, características apresentadas anteriormente como pontos de estresse em arquiteturas centralizadas. O domínio piloto selecionado pode ser expandido para mais de um, desde que a equipe designada para o projeto tenha a capacidade necessária para absorver essa expansão sem comprometer o planejamento estabelecido.

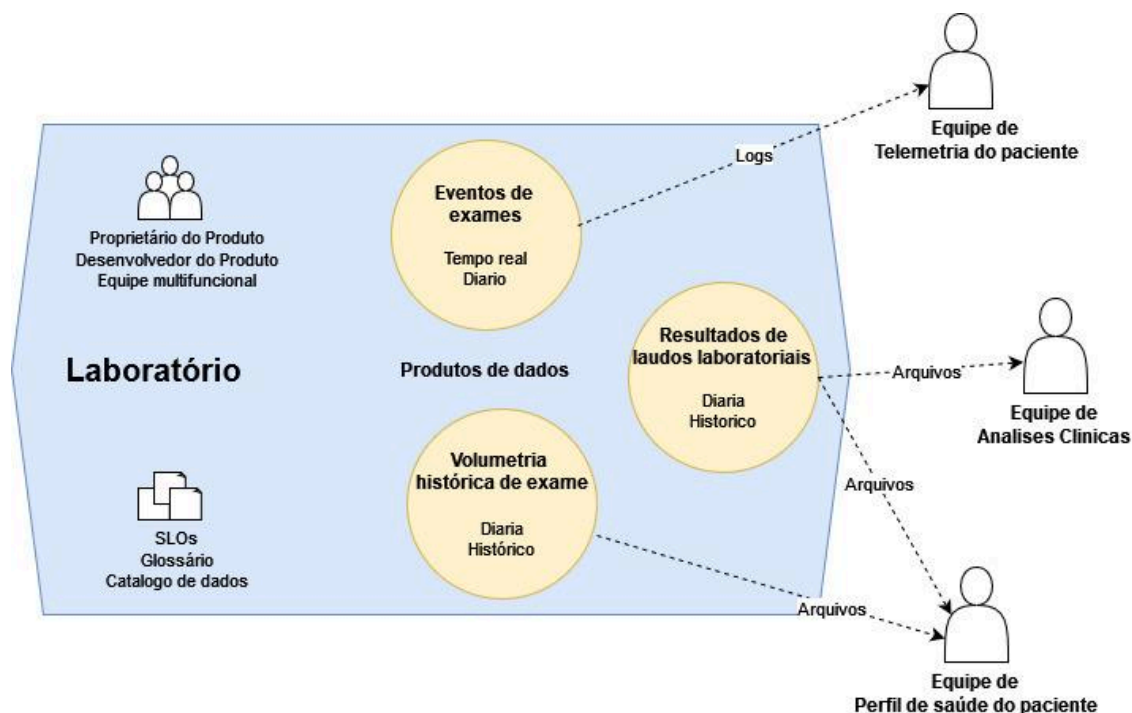
### **3.3.1 Definição de um produto de dados**

Em decorrência da possibilidade de criação de silos ao descentralizarmos os domínios de dados da instituição, os produtos de dados surgem como uma abordagem para mitigar esse problema (DEHGHANI, 2023). Eles promovem uma mudança de perspectiva, tratando os dados não mais como meros ativos, mas como produtos com valor intrínseco e propósito claro. Um produto de dados pode ser definido como um conjunto de informações que possui características específicas, tornando-se valioso para os usuários (DEHGHANI, 2023). Essas características incluem: capacidade de descoberta, endereçamento, compreensão, confiabilidade, acessibilidade nativa, interoperabilidade, valor independente e segurança (DEHGHANI, 2023). Esses atributos ajudam a diferenciar os produtos oferecidos por cada domínio, transformando-os em elementos vivos, projetados para atender às necessidades dos usuários e garantir sua confiança no que lhe é entregue. Para que essa abordagem seja bem-sucedida, é necessário promover uma mudança cultural e estratégica dentro da instituição, alinhando pessoas, processos e tecnologias para assegurar que os produtos de dados sejam sustentáveis.

### **3.3.2 Identificando os produtos de dados do domínio piloto: Laboratório**

Ao compreender os conceitos que definem um produto de dados, a Figura 5 apresenta uma proposta detalhada dos principais produtos identificados no domínio inicial de laboratório, destacando seus consumidores e suas interações. Essa proposta foi elaborada com base em produtos de dados reais observados em um setor de laboratório de um hospital, refletindo os fluxos informacionais comuns dessa área. A estrutura proposta considera o ambiente centralizado atual, ilustrado na Figura 3, e a distribuição dos domínios conforme descrito na Figura 4, servindo como um ponto de partida para a transição gradual rumo à descentralização.

Figura 5 – Definição dos produtos de dados do domínio de laboratório.



fonte: De autoria própria

Podemos verificar que existem diversas necessidades e para conseguir atender a todos, é disponibilizado produtos diversos, com características diferentes, por exemplo, o produto de dados que captura eventos de exames. Esse produto transmite em tempo real, via streaming, uma série de logs que informam quais exames estão sendo realizados, o equipamento utilizado, o local em uso, entre outros dados relevantes. Esses dados são essenciais para que a equipe de telemetria do paciente obtenha informações confiáveis, permitindo o mapeamento dos pacientes na instituição.

Além disso, é disponibilizado um outro produto voltado para resultados de laudos laboratoriais, que converte informações desestruturadas de PDFs para arquivos estruturados, disponibilizando-os diariamente em repositórios históricos. Esses dados estruturados permitem que outras equipes acessem essas informações de maneira eficiente para efetuar suas análises.

Assim, é possível verificar que a equipe multifuncional do domínio de laboratório é composta por especialistas que possuem conhecimento aprofundado sobre o

negócio, entendem a estrutura dos sistemas de dados, estão familiarizados com a formatação dos laudos em PDF e dominam a terminologia e os conceitos médicos aplicados, garantindo maior qualidade, agilidade e eficiência no desenvolvimento dos produtos de dados desse domínio.

### **3.3.3 Definição da equipe multifuncional e seus papéis no domínio piloto**

A equipe multifuncional, conforme apresentada, pode incluir diversos profissionais, com especializações que variam de acordo com a complexidade e as necessidades específicas de cada domínio. Para o domínio de laboratório, as possíveis funções mapeadas para essa equipe, destacam-se médicos especialistas, engenheiros de dados, cientistas de dados e analistas de negócios e sistemas, bem como, dois outros papéis essenciais, sendo eles: o proprietário do produto de dados e o desenvolvedor de produtos de dados.

O proprietário do produto de dados é alguém com profundo conhecimento dos dados abrangentes do domínio, enquanto o desenvolvedor de produtos de dados é responsável por gerenciar e criar os produtos de dados do domínio (DEHGHANI, 2023). Ambos têm como principais responsabilidades assegurar a confiabilidade dos produtos entregues, além de ter autoridade para tomar decisões internas durante o desenvolvimento. Esses papéis podem ser desempenhados por uma única pessoa ou por um grupo de profissionais, dependendo da estrutura e das demandas do domínio.

### **3.3.4 Gestão dos produtos e dados centrais da instituição**

Observa-se que um domínio pode conter múltiplos produtos de dados e atender a uma diversidade de usuários. Nesse contexto, é essencial que cada produto de dados defina claramente seu escopo de entrega, o método de disponibilização, a periodicidade, os mecanismos de acesso e as instruções de uso. Para garantir esse objetivo, é necessário disponibilizar documentos que apoiem o entendimento e a utilização dos dados, como glossários, *Service Level Objectives* (SLOs), catálogos entre outros (DEHGHANI, 2023). Além disso, deve-se designar responsáveis, como o proprietário do domínio, para orientar e ajudar os usuários quanto ao uso

adequado dos dados, bem como, a criação de uma comunidade para estabelecer uma cultura voltada a dados. (DPG, 2023).

Esse processo requer uma governança de dados eficiente, tanto dentro do domínio quanto entre diferentes domínios de dados, especialmente para dados centrais, como os de pacientes, médicos e unidades. Esses dados são fundamentais para garantir a rastreabilidade dos pacientes e sua integração entre os domínios, de modo que seja necessário estabelecer padrões que contemplem essas necessidades, respeitando a perspectiva analítica de cada domínio. A Governança Federada, será discutida em maior profundidade na seção 3.5 deste trabalho.

### **3.4 Elaboração do planejamento estratégico de uma plataforma de autoatendimento**

Um dos principais desafios para a implementação dos conceitos discutidos anteriormente, como a criação de um ambiente descentralizado em domínios e a proposta de uma arquitetura distribuída via produtos de dados, está na complexidade do desenvolvimento de uma plataforma de dados que viabilize essa transformação. O aumento da autonomia dos domínios levanta questões sobre os custos associados ao desenvolvimento e seu potencial retrabalho envolvido em todo o processo de construção (DEHGHANI, 2023).

Assim, o objetivo deste capítulo é apresentar as características e os requisitos que uma plataforma de autoatendimento deve atender, destacando suas particularidades em relação às plataformas de dados tradicionais; oferecendo uma proposta de direcionamento estratégico a gestores e tomadores de decisão em como implementar tais plataformas. Iniciando o processo com base no domínio de partida definido na seção anterior. Por ser um direcionamento estratégico a análise desenvolvida não irá se adentrar em ferramentas ou softwares específicos. Abordando uma perspectiva macro sobre os pontos discutidos, pois uma análise detalhada e técnica de uma implementação micro foge ao escopo desta monografia, sendo mais apropriada para estudos dedicados exclusivamente a essa abordagem operacional.

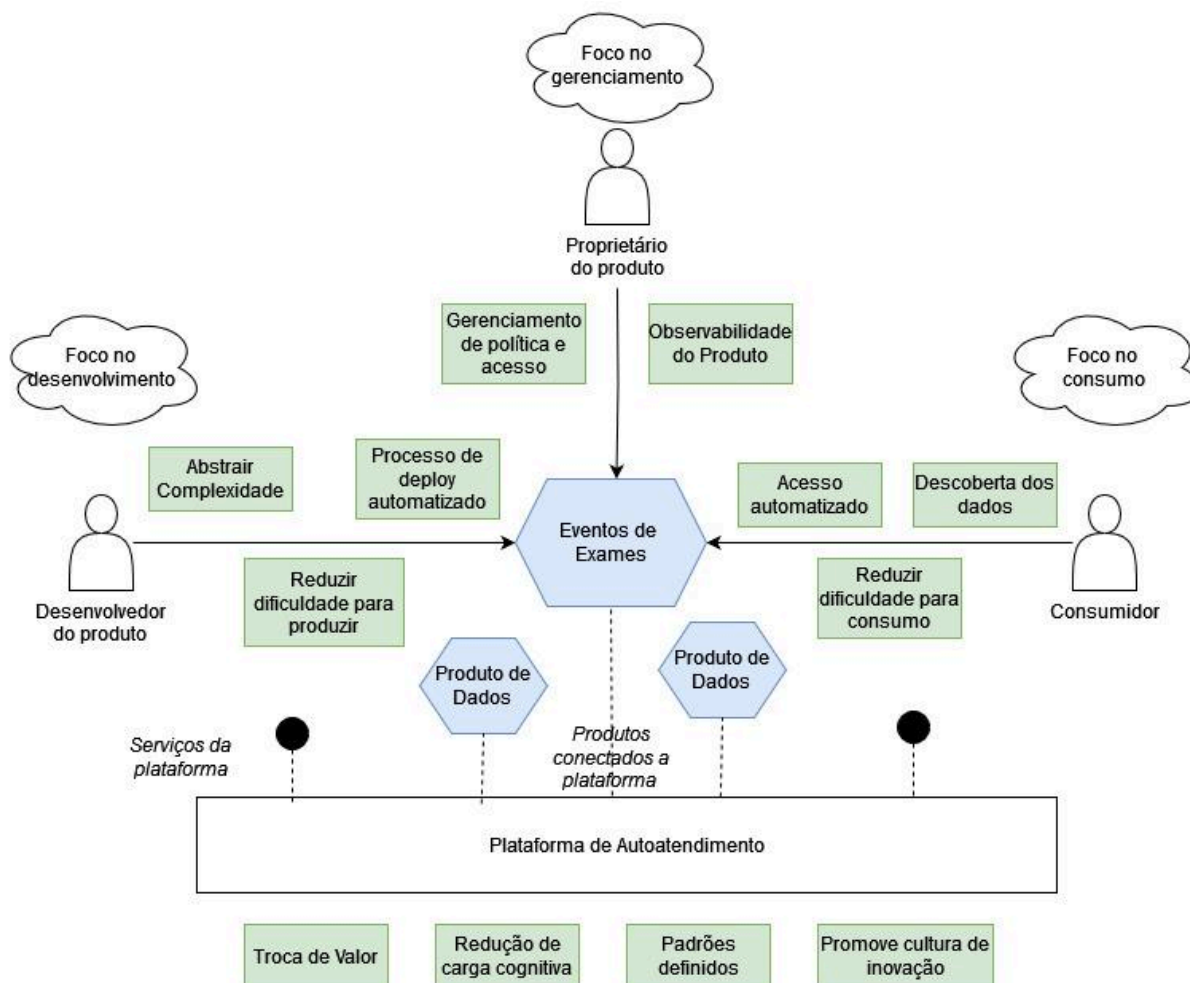
### 3.4.1 Definição e características de uma plataforma de autoatendimento

Podemos definir uma plataforma como um local onde simplificamos e automatizamos tecnologias complexas e reduzimos a carga cognitiva para sua utilização (SKELTON; PAIS., 2019), dessa forma, é através dela que abstrairmos todas as complexidades relacionadas a produção e consumo dos dados e buscamos automatizar grande parte do trabalho operacional necessário para manter, gerenciar e implementar uma solução analítica, tornando possível a utilização de forma eficiente pela instituição.

Uma plataforma de autoatendimento voltada para data mesh em saúde não difere dessa definição, é através dela que será possível a implementação de produtos de dados de forma eficiente e aplicação das políticas definidas pela instituição de forma automatizada, porém, ao contrário das plataforma atuais onde separamos o código, a política e os dados, uma plataforma de autoatendimento trata um produto de dados em sua totalidade de forma independente (DEHGhani, 2023), ou seja, o produto contém desde o armazenamento a questões de segurança, se tornando autônomo nesse ecossistema. Assumindo uma postura descentralizada, como uma rede interconectada de produtos de dados se comunicando entre si, conseqüentemente elevando a complexidade para sua implementação (DEHGhani, 2019).

Dessa forma, uma plataforma descentralizada apresenta características específicas para viabilizar a implementação deste paradigma sociotécnico na organização. Na Figura 6, é destacado as principais características e expectativas descritas pela autora Dehghani (2023), que orientam a construção de um ambiente descentralizado. Esses conceitos foram aplicados ao cenário centralizado real em saúde ilustrado na Figura 3 e ao domínio piloto e seus produtos de dados ilustrados na figura 5, fundamentais para iniciar a transição de um modelo centralizado para um distribuído.

Figura 6 – Características de uma plataforma de autoatendimento



fonte: De autoria própria

Conforme visto na figura acima as característica de uma plataforma voltada para data mesh em saúde, leva em consideração o produto de dados como um todo (armazenamento, dado, código, segurança e etc) e como esse produto se relaciona com os usuários e outros produtos dentro desse ecossistema. Conforme visto no capítulo 3.3.1, um produto deve atender às expectativas de quem o utiliza, logo, a plataforma é um produto interno que será utilizado pelo produtor, gerenciador e consumidor de dados e precisa atender suas necessidades (DEGHANI, 2023).

Essa plataforma, portanto, deve estabelecer os padrões de arquitetura e compartilhamento a serem cumpridos por cada produto de dados, o que garantirá a interoperabilidade entre os domínios consequentemente aumentando a troca de valor entre eles, ao mesmo tempo que promove uma cultura de inovação, reduzindo

a carga cognitiva dos usuários ao abstrair as complexidades de seu uso, criando assim uma plataforma mais generalista. Isso tende a aumentar o número de usuários e a impulsionar o uso de dados dentro da instituição. (DEHGHANI, 2019).

Sob a ótica do consumidor, um dos principais usuários da plataforma, espera-se que seja possível realizar a descoberta dos produtos de dados disponíveis na instituição de maneira transparente, permitindo ao usuário compreender todas as informações essenciais para sua utilização, principalmente na área da saúde, no qual apresenta siglas e termos médicos que precisam ser compreendidos com exatidão. O acesso a esses produtos deve ser configurado de forma automatizada, baseado em políticas e perfis de acesso, reservando aprovações manuais para exceções e não como regra. Essa abordagem visa reduzir a complexidade do consumo, descoberta e compreensão dos produtos de dados no ambiente.

Na perspectiva dos desenvolvedores responsáveis pela produção e edição de produtos de dados, o foco principal reside na redução da complexidade dessas tarefas. É esperado que as operações sejam abstraídas de modo a permitir que os desenvolvedores concentrem-se exclusivamente nas etapas de desenvolvimento, teste, validação e publicação do produto, sem a necessidade de se preocupar com questões arquiteturais do ambiente. Efetuando o processo de implantação de forma automatizada, com mecanismos que permitam a entrega do produto em sua totalidade (código, políticas de acesso, documentação de uso, dados e glossário), garantindo uma maior velocidade para publicar novos produtos de dados padronizados.

Por fim, sob a ótica do produtor de dados, o objetivo principal é gerenciar esse ambiente de forma simplificada, mantendo uma visibilidade ampla sobre os processos que ocorrem dentro de cada produto de dados e do ambiente como um todo, bem como, assegurando a conformidade com as políticas e normas estabelecidas para o domínio e instituição. A plataforma deve oferecer mecanismos de observabilidade que permitam ao produtor monitorar a qualidade dos dados entregues e garantir que os produtos estejam em conformidade com os requisitos institucionais e regulatórios, Abordaremos com mais detalhes os requisitos de

gerenciamento no t3pico 3.5.2 Requisitos para aplicar o modelo de governan3a em data mesh na sa3de.

### 3.4.2 Elab3r3o da estrat3gia para implementa3o de uma plataforma de autoatendimento em sa3de

Uma proposta estrat3gica desenvolvida e implementada em unidades hospitalares da am3rica latina para introduzir as plataformas de autoatendimento ao mesmo tempo que mant3m sua plataforma atual de dados, se baseia em entregas graduais, atacando pontos espec3ficos das caracter3sticas comentadas acima a partir de um dom3nio piloto e evoluindo a plataforma aos poucos atrav3s de um planejamento bem definido, visto que, novas cl3nicas, seguradoras e hospitais s3o credenciados e descredenciados dinamicamente, tornando esse ecossistema din3mico. Dessa maneira a tabela 1 apresenta um planejamento detalhado e ordenado para guiar a estrat3gia nesse processo de constru3o.

Tabela 1 – Proposta de planejamento para implementar uma plataforma data mesh em sa3de.

(continua)

Etapa	Descri3o
1 - Definir um dom3nio de partida	Conforme discutido anteriormente, o dom3nio inicial permite a constru3o de um ambiente reduzido, buscando uma prototipa3o mais r3pida, com o objetivo de solucionar problemas de forma antecipada, al3m de diminuir a complexidade envolvida no desenvolvimento de uma plataforma completa. No contexto deste planejamento, o dom3nio selecionado como ponto de partida foi o Laborat3rio.
2 - Definir/Revisar ferramentas atuais	Ao analisar o ecossistema atual e as ferramentas de desenvolvimento dispon3veis, 3 poss3vel identificar o grau de ader3ncia dessas tecnologias 3s caracter3sticas de uma plataforma de autoatendimento, conforme discutido anteriormente. Essa avalia3o possibilita um planejamento mais eficiente, reduzindo os custos e otimizando o tempo necess3rio para a implanta3o. Aproveitando o que j3 foi criado e 3 utilizado pela institui3o.

Tabela 1 – Proposta de planejamento para implementar uma plataforma data mesh em saúde.

(continuação)

Etapa	Descrição
3 - Definir/Evoluir os padrões da plataforma	Essa etapa é fundamental para assegurar a interoperabilidade dos produtos de dados e viabilizar a construção de uma plataforma descentralizada. Para isso, é necessário definir os protocolos e os modelos de APIs que serão utilizados na comunicação, estabelecer padrões para a disponibilização e organização dos arquivos, detalhar as possibilidades de conexão entre os sistemas, entre outros pontos. Além disso, é crucial que essas definições sejam seguidas e replicadas de forma consistente em cada produto de dados, garantindo uma uniformidade no ecossistema como um todo respeitando as particularidades de cada um.
4 - Criar/Melhorar a plataforma de consumo	Sob a perspectiva do consumidor e considerando suas necessidades, um ponto de partida estratégico é a criação e a evolução de uma plataforma para o consumo de produtos de dados. Essa plataforma deve simplificar o acesso à informação, promovendo a descoberta intuitiva dos produtos e permitindo a conexão e utilização de forma independente deles. Com isso é possível, tanto o consumo de novos produtos de dados quanto o acesso aos dados já disponíveis na plataforma centralizada existente, facilitando a transição para uma arquitetura descentralizada.
5 - Desenvolver um produto de dados em sua totalidade	Essa etapa envolve o desenvolvimento de um produto de dados mapeado no domínio piloto, com o objetivo de reduzir a complexidade associada à criação e edição deste produto. Além disso, busca-se automatizar os processos de desenvolvimento, teste e publicação, garantindo a entrega do produto em sua totalidade dentro do ecossistema de forma independente, ponto este primordial. Esse produto deve estar em conformidade com os padrões previamente estabelecidos e respeitar as políticas definidas, assegurando sua interoperabilidade e disponibilidade para consumo como uma rede interconectada.

Tabela 1 – Proposta de planejamento para implementar uma plataforma data mesh em saúde.

(conclusão)

Etapa	Descrição
6 - Monitorar os produtos de dados do ambiente	Essa etapa busca garantir a observabilidade dos processos internos do produto de dados e do ambiente como um todo, para assegurar transparência e controle. Isso é possibilitado por meio de APIs e gráficos que apresentem informações de forma clara. Além de facilitar o monitoramento da qualidade do produto, essa abordagem permite verificar sua conformidade com as políticas e normas estabelecidas pelo domínio e verificar duplicidades e inconsistência no ambiente, promovendo uma gestão alinhada aos contratos definidos.
7 - Aprendizado contínuo e adesão de novos produtos e domínios voltando a etapa 1	É fundamental identificar, mapear e documentar os conflitos e problemas encontrados durante o desenvolvimento, bem como suas respectivas soluções, promovendo o aprendizado a partir de produtos previamente implementados. Esse processo deve ser cíclico, introduzindo novos produtos na plataforma com base em um planejamento bem estruturado. Após a finalização de um domínio, um novo é selecionado para desenvolvimento, até que a plataforma atinja um nível de maturidade suficiente para viabilizar uma migração completa do ambiente de forma segura

Fonte: De autoria própria

Conforme previamente mencionado, existem diversas ferramentas disponíveis que permitem a implementação de uma plataforma descentralizada. Essas ferramentas oferecem funcionalidades como catálogos, linhagem de dados, integração com diversos sistemas, capacidade de consumir e relacionar diferentes conjuntos de dados, entre outros. Contudo, o foco deste estudo está na estratégia de introdução de uma plataforma, não nos detalhes técnicos de sua implementação, visto que, a escolha de ferramentas envolve uma série de outros fatores que precisam ser analisados pela organização, abrindo possibilidades de pesquisas focadas nesse tema.

Ao seguir o planejamento e adaptá-lo à sua realidade é possível efetuar a migração da plataforma existente ao mesmo tempo que o conceito de data mesh e sua cultura se dissemina na instituição de forma natural, sendo uma plataforma viva no qual evolui e engloba mais produtos, além de aumentar seu uso, buscando escalar mais facilmente que o modelo centralizado atual e diminuir o tempo que uma solução analítica é idealizada até sua publicação, sem os problemas mencionados anteriormente nesta monografia. Respeitando as particularidades do segmento de saúde, uma vez que cada produto é independente e processa os dados de maneira específica para atender às suas necessidades.

### **3.5 Modelo de governança federada**

Conforme identificado nessa etapa da monografia, a complexidade para gerenciamento e monitoramento de uma plataforma de autoatendimento descentralizada que se assemelha a uma rede interconectada de produtos de dados e domínios é elevada consideravelmente, exigindo que a governança atual de dados se adapte a essa nova realidade, assumindo uma postura de orientação e não mandatória e federando seu controle, com gerenciamento e políticas internas dos domínios, bem como, políticas e padrões globais do ambiente, além das tarefas já efetuadas de qualidade, monitoramento e entre outros. Sendo um ponto crucial para diminuir os possíveis problemas enfrentados na descentralização e sucesso da implementação do paradigma sociotécnico do data mesh (DEHGHANI, 2023).

Assim este tópico tem como objetivo apresentar os principais princípios que regem uma governança federada em um ambiente descentralizado e seus requisitos, destacando os possíveis desafios a serem solucionados e propondo uma abordagem estratégica para a aplicação desse novo modelo em um ambiente descentralizado em saúde.

#### **3.5.1 Definição de governança federada e suas diferenças para a governança de dados tradicional.**

Podemos definir a governança de dados como uma abordagem com princípios de qualidade, segurança e integridade dos dados para gerenciar esse ativo durante

todo o ciclo de vida, desde a aquisição até o uso e o descarte (GOOGLE, 2024b). Uma governança de dados eficiente reduz o custo de operação de uma plataforma de dados, aumenta a qualidade e confiabilidade desse ativo na instituição e viabiliza que soluções analíticas gerem impactos positivos, resultando em decisões mais assertivas. Além de promover uma cultura organizacional orientada a dados, sendo um dos pilares determinantes para o sucesso na implementação de soluções e manutenção da competitividade da instituição no mercado. (GOOGLE, 2024b).

Esse objetivo se mantém ao abordarmos uma governança em um ambiente descentralizado, No entanto, a complexidade do gerenciamento nesse cenário exige adaptações que tornem esse papel federado entre os proprietários de produtos de dados apresentado na seção 3.3.3; representantes da plataforma e equipes de segurança e jurídico. Dessa forma, os domínios de dados mantêm autonomia e poder de decisão local, aplicando normas, políticas internas e definições de métricas específicas de qualidade que respeitem suas particularidades, ao mesmo tempo em que seguem padrões e políticas globais aplicáveis a todo o ambiente (DEHGHANI, 2023).

As regras globais são definidas de maneira colaborativa, envolvendo os proprietários de produtos de cada domínio, os representantes da plataforma, e as equipes jurídica e de segurança. Essa abordagem garante tanto a conformidade legal quanto a observabilidade de todo o ambiente, alinhando a governança aos objetivos estratégicos do negócio; diminuindo a possibilidade de criação de silos e viabilizando a construção de uma plataforma de autoatendimento funcional. (DEHGHANI, 2023).

### **3.5.2 Requisitos para aplicar o modelo de governança em data mesh na saúde**

Além da aplicação do modelo federado, explicado acima, e os objetivos já efetuados da governança de dados tradicional, o modelo de governança em data mesh apresenta outros 2 pilares que precisam ser introduzidos para a minimização de problemas da descentralização, principalmente no cenário em saúde, com dados sensíveis e pessoais em sua grande maioria, sendo eles o pensamento sistêmico e a execução computacional. Esses pilares alteram a percepção em como os dados

são gerenciados, determinando os requisitos e características que precisam ser introduzidas neste novo modelo e na plataforma de autoatendimento comentada acima. Abaixo definimos suas particularidades segundo a autora Dehghani (2023).

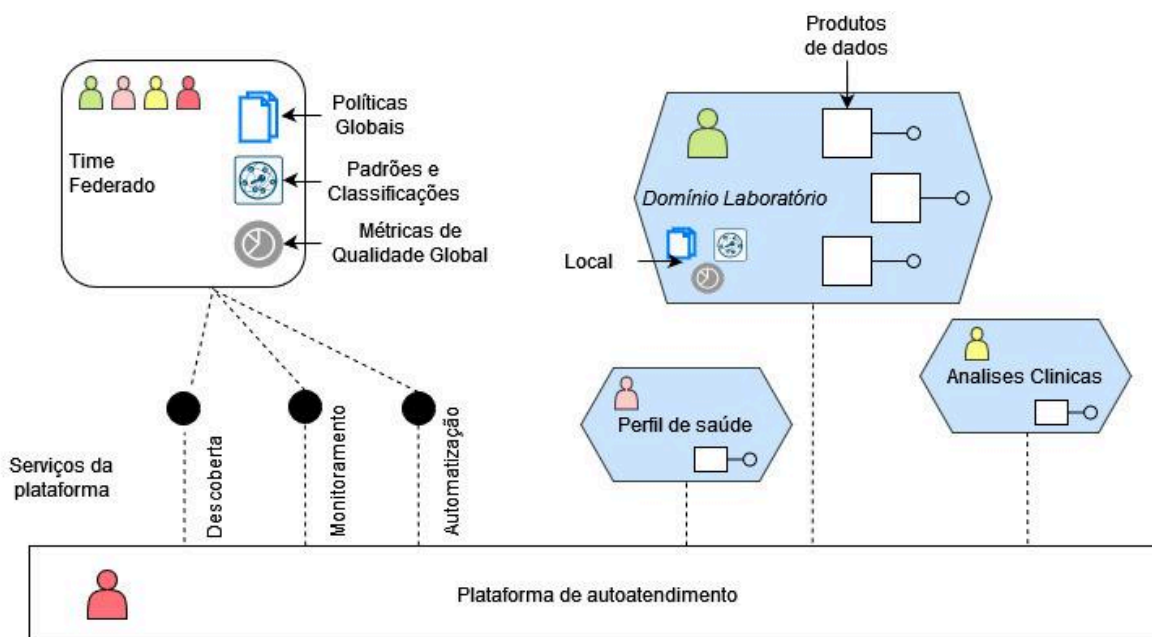
O pensamento sistêmico propõe uma visão holística e integrada dos produtos de dados, tratando-os como parte de um sistema interconectado em vez de ver os produtos como elementos isolados. Identificando interações entre os domínios e como isso afeta o ambiente, sendo necessário a criação de padronizações e políticas globais para garantir a interoperabilidade dos domínios, bem como, a observabilidade com métricas de qualidade, uso, entre outras para avaliação de duplicações e inconsistência dos produtos entregues, efetuando feedbacks contínuos e correções necessárias para manter um ambiente saudável.

A execução computacional envolve a integração de práticas de governança diretamente nos sistemas e processos de dados, aumentando a automação e o uso de infraestrutura como código (IaC) na plataforma de autoatendimento. Assegurando assim que as políticas de dados, segurança e conformidade sejam aplicadas de forma escalável, promovendo uma eficiência operacional ao reduzir intervenções manuais e permitir uma gestão mais proativa dos dados, garantindo que as regras de governança sejam seguidas automaticamente em toda a organização.

### **3.5.3 Implantação de uma governança federada ao ambiente descentralizado**

A partir dos requisitos identificados anteriormente e com a construção de uma plataforma que possibilite minimamente a observabilidade e automatização do ecossistema é possível aplicarmos esse novo modelo no ambiente descentralizado apresentado neste capítulo. Na figura 7 representamos uma visão macro do modelo de governança federada em data mesh aplicados nos domínios e plataforma desenvolvidas nas seções anteriores.

Figura 7 – Visão macro da governança federada aplicada ao ambiente



fonte: De autoria própria

Ao analisar a figura acima, podemos observar a aplicação do novo modelo, que segue os princípios previamente apresentados. Nesse contexto, um time federado, composto por múltiplos representantes do ambiente, é responsável pela definição de padrões, métricas e políticas globais. Aplicado ao domínio laboratório, isso envolve, entre outros aspectos, a definição de métricas de qualidade relacionadas à quantidade de dados nulos e à presença de datas inconsistentes nos produtos, o tempo de resposta dos logs em tempo real e os padrões de formatação para dados pessoais de pacientes. Isso se torna possível graças aos serviços da plataforma de autoatendimento, que facilita a descoberta de novos produtos de dados, efetua o monitoramento dos SLOs de entrega e automatiza os processos para aplicação das políticas e normas definidas. Dessa forma, é possível manter uma visão holística do ambiente, mesmo que ele esteja descentralizado.

Os domínios de dados seguem as regras e padrões globais estabelecidos, mas também possuem autonomia para contribuir na criação e edição dessas normas. Além disso, são responsáveis pela elaboração de políticas e métricas internas para cada produto entregue, respeitando suas características específicas. Exemplos aplicados no domínio de laboratório incluem a criação de métricas de satisfação do

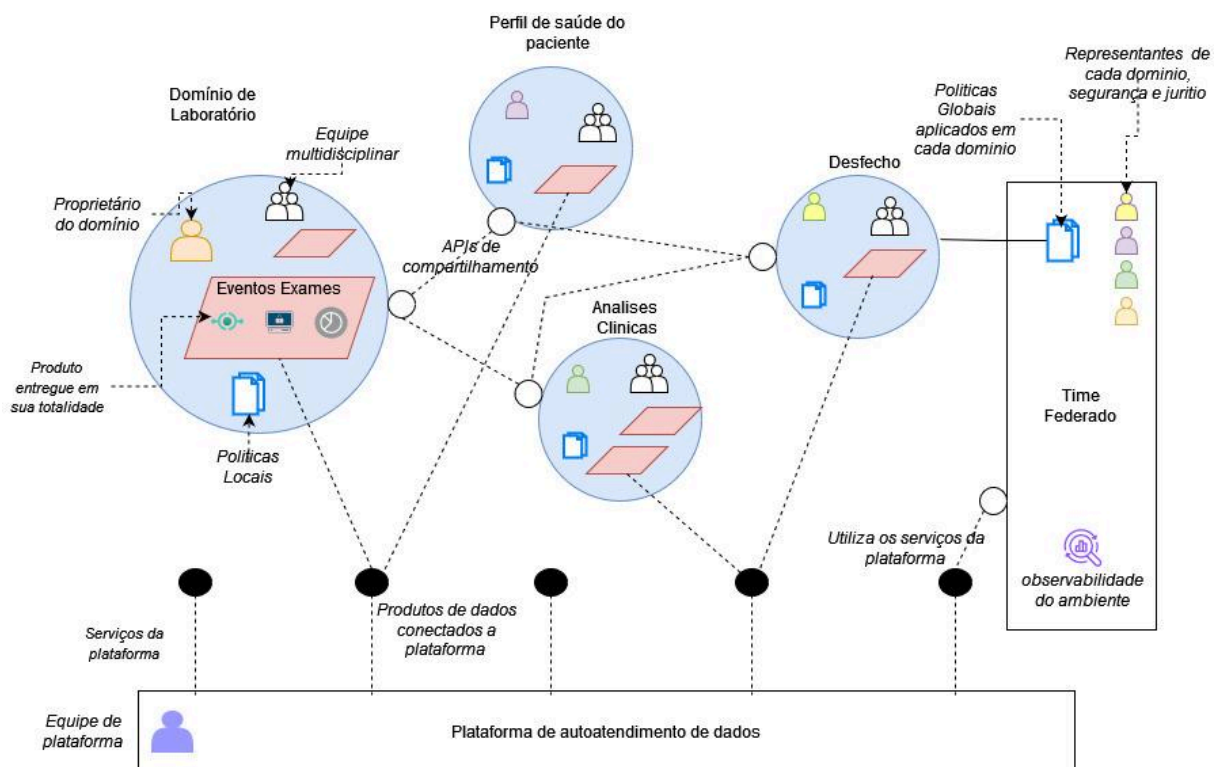
usuário, que podem ser mais específicas do que as instituídas em nível organizacional, bem como a aplicação de normas ou legislações específicas, como o mascaramento e anonimização de dados pessoais dos exames laboratoriais. Esse processo garante uma maior aderência a normas e regulamentações, especialmente no contexto da saúde, possibilitando a implementação de um modelo descentralizado funcional.

### **3.6 Ambiente informacional descentralizado: Conclusão**

Ao aplicarmos os princípios apresentados durante o todo esse capítulo, obtemos um sistema descentralizado de dados no qual busca disseminar uma cultura voltada a dados, focando em aumentar a autonomia dos domínios de dados, tornando os produtores e não apenas consumidores além de capacitá los para isso, bem como, aproximando os especialista do negócio com os produtos que eles geram. Outro aspecto é o aumento da escalabilidade da plataforma, replicando padrões de arquitetura e comunicação para introduzir novos produtos de dados como uma rede interconectada de maneira eficaz. E por fim federando a governança de dados, permitindo uma visão holística do ambiente e monitoramento da adesão a normas e políticas dos domínios e entre os domínios.

Tudo isso é possível graças a uma mudança organizacional, estratégica e cultural na organização, alterando a percepção em como o ativo de dados é visto, consumido e apresentado, baseados nos princípios que regem a proposta sociotécnica do data mesh e os adaptando ao setor de saúde com suas particularidades. Na Figura 8, resumimos todos os conceitos já previamente apresentados, com uma visão macro do ambiente descentralizado desenvolvido no decorrer deste capítulo.

Figura 8 – Ambiente informacional descentralizado final com data mesh



fonte: De autoria própria

## 4. ANÁLISE DE RESULTADOS

Para avaliar a viabilidade da descentralização fundamentada no princípio sociotécnico do data mesh, este capítulo apresenta uma análise comparativa entre o ambiente centralizado inicial e o ambiente descentralizado proposto, buscando examinar o comportamento de cada configuração, verificando se os problemas identificados no modelo centralizado persistem no novo ambiente. Além disso, são apresentados estudos de caso de instituições que estão implementando o modelo descentralizado, destacando os principais benefícios alcançados, os desafios enfrentados e os aprendizados mapeados ao longo do processo.

### 4.1 Comparação do ambiente centralizado inicial e do ambiente descentralizado gerado

Após analisar os ambientes ilustrados nas Figuras 3 e 8, a Tabela 2 apresenta uma comparação dos principais aspectos de um ambiente informacional, destacando como cada um deles se comporta no contexto de um ambiente centralizado em contraste com um ambiente descentralizado.

Tabela 2 – Comparação dos ambientes centralizado e descentralizado

(continua)

Aspecto	Ambiente centralizado	Ambiente descentralizado (data mesh)
Ingestão de Dados	Os dados são ingeridos em um único ponto centralizado (data lake).  - Processos de ETL centralizados.	Os dados são ingeridos nos domínios de negócio, próximos à fonte de dados.  - Cada domínio é responsável pela ingestão e transformação de seus próprios dados.

Tabela 2 – Comparação dos ambientes centralizado e descentralizado

(conclusão)

Aspecto	Ambiente centralizado	Ambiente descentralizado (data mesh)
Transformação de Dados	<p>A transformação é realizada em um ponto centralizado, com governança central.</p> <p>- Modelos de dados globais são definidos.</p>	<p>Cada domínio de dados realiza suas próprias transformações, alinhadas às necessidades específicas do domínio.</p> <p>- Governança descentralizada com padrões globais definidos para interoperabilidade.</p>
Consumo de Dados	<p>Os dados são acessados centralizadamente, com visualizações em plataformas centralizadas.</p> <p>- Consumo dependente da equipe central.</p>	<p>O consumo de dados é descentralizado, cada domínio expõe seus dados para consumo com APIs, data products ou services.</p> <p>- Consumidores podem acessar dados diretamente de cada domínio.</p>
Governança	<p>Governança centralizada, com uma equipe responsável por todas as políticas e controle de qualidade.</p> <p>- Definição de padrões e normas para os dados em toda a organização.</p>	<p>Governança descentralizada com padrões comuns (como modelos de dados e APIs).</p> <p>- Cada domínio é responsável pela governança de seus próprios dados, mas segue diretrizes globais para garantir a interoperabilidade.</p>
Monitoramento e Observabilidade	<p>Monitoramento centralizado dos processos de ingestão, transformação e consumo dos dados.</p> <p>- Métricas e logs são coletados em um sistema centralizado.</p>	<p>Monitoramento descentralizado nos domínios, com responsabilidade dos times de dados de cada domínio.</p> <p>- Ferramentas de observabilidade como métricas e logs são implementadas em cada domínio, com relatórios para visibilidade global</p>

Fonte: De autoria própria

A análise da tabela acima evidencia que os desafios da centralização em sistemas de saúde apresentados nesta monografia são minimizados com a adoção da estratégia de data mesh. Problemas iniciais, como o distanciamento entre as equipes produtoras de dados e a proliferação descontrolada dos dados, são reduzidos pela criação de domínios de dados que atuam tanto como produtores quanto como consumidores. Além disso, a dificuldade de escalabilidade do ambiente centralizado, é superada pela replicação de padrões e modelos de comunicação, permitindo a adição de novos produtos à plataforma de forma estruturada, com a disponibilização de APIs e serviços a rede interconectada criada.

Outro benefício é a aplicação uniforme de regulamentações e normas específicas do setor de saúde, como as exigências da LGPD. Essas regras são incorporadas tanto no ambiente global quanto nos domínios específicos, assegurando conformidade legal e promovendo um monitoramento mais detalhado em produtos que demandam maior atenção.

#### **4.2 Cases de sucesso na implantação do data mesh**

A descentralização de ambientes informacionais de dados por meio da abordagem do data mesh não se limita ao setor de saúde. Outros segmentos também têm explorado a aplicação desses conceitos, especialmente quando enfrentam desafios semelhantes aos descritos neste trabalho, como a existência de domínios de dados densos e em constante expansão, a presença de múltiplas partes interessadas e a diversidade de sistemas geradores de dados, encontrando através dos princípios do data mesh uma abordagem para construir uma organização orientada a dados, promovendo a disseminação das responsabilidades de criação e gestão entre diferentes áreas da instituição, além de fomentar a capacitação e letramento sobre soluções analíticas.

Nesse cenário, é possível identificar cases de sucesso de empresas que têm adotado esse paradigma sociotécnico, mesmo sendo uma abordagem relativamente recente. Muitos desses casos já oferecem insights valiosos sobre os benefícios e os desafios dessa transformação organizacional.

#### **4.2.1 Implementação dos princípios do data mesh na DPG Media**

Conforme visto no artigo escrito pela DPG Media It (2021) a empresa adotou os conceitos e princípios da descentralização do data mesh buscando se tornar verdadeiramente orientada a dados. Em 2021, os dados estavam no centro de sua estratégia digital, mas a abordagem centralizada apresenta desafios, como gargalos causados por equipes limitadas e dificuldades em priorizar demandas. O data mesh foi visto como uma solução para permitir a cultura orientada a dados em toda a organização, descentralizando responsabilidades e capacitando mais equipes para criar produtos de dados com autonomia.

Inicialmente, em 2019, centralizaram esforços para resolver problemas de governança e organizar dados em quatro domínios principais, colocando engenheiros de dados junto aos especialistas do domínio, como no caso do sistema de assinaturas, gerando aprendizado e autonomia. Embora desafiador, o teste mostrou o valor da abordagem descentralizada para expansão futura do data mesh.

A DPG Media planeja expandir sua visão descentralizada de data mesh em 2022. O foco está em clarificar impactos orçamentários, implementar governança e qualidade de dados e identificar proprietários de produtos de dados. A empresa busca construir comunidades de dados para compartilhar conhecimento técnico e apoiar departamentos com ferramentas e governança. A estratégia também inclui estruturar guildas técnicas, fomentar especialistas em ferramentas de autosserviço e criar grupos de governança de dados com perspectivas técnicas e operacionais. A empresa enxerga um futuro promissor nessa abordagem e estima que em poucos anos, toda a organização estará descentralizada.

#### **4.2.2 Implementação dos princípios do data mesh na Globo**

Conforme apresentado no vídeo disponibilizado pela Globo (2021), Éfrem de Aguiar desenvolvedor do Hub Digital da empresa comenta que a instituição identificou limitações no modelo centralizado de gestão de dados, como dificuldades de escalabilidade e agilidade no acesso às informações. Esses desafios

comprometeram a capacidade da organização de atender à crescente demanda por dados rápidos em um ambiente cada vez mais dinâmico.

Para superar essas limitações, a Globo implementou o data mesh, estruturando seus dados em domínios autônomos, com equipes responsáveis pela criação e gestão de seus próprios conjuntos de dados. A governança foi reforçada por contratos que garantem a consistência e o compartilhamento adequado de dados entre as equipes. Além disso, a infraestrutura foi construída utilizando ferramentas avançadas da *Google Cloud Platform* (GCP), como o *Dataform* e o *Composer*, adaptando também ferramentas internas para garantir monitoramento contínuo, atualização de dados e acesso nativo.

A transformação levou a um investimento de cerca de R\$ 5 milhões em inovação e produtos de dados, resultando em melhorias significativas na eficiência operacional e na entrega de insights. O sistema alcançou um *uptime* de 99%, garantindo a disponibilidade dos dados para processos críticos, com impacto direto ou indireto para mais de 100 milhões de usuários. Além disso, a adoção do data mesh possibilitou à Globo a expansão de sua capacidade analítica e o fortalecimento da cultura de dados, com foco na autonomia e colaboração das equipes.

### **4.3 Considerações do Capítulo**

Como podemos avaliar a abordagem dos princípios do data mesh, quando justificada por motivos claro da limitação da centralização do ambiente e da necessidade de expansão de maneira descentralizada da instituição, pode gerar resultados promissores, apesar de seus desafios para implementação, sua mudança cultural na instituição e a autonomia dos domínios de dados permite a criação de um ambiente colaborativo, reduzindo muito a exaustão operacional de equipes centrais e melhorando as análises geradas com dados confiáveis.

## **5. CONSIDERAÇÕES FINAIS**

Com a adoção de uma proposta descentralizada em instituições de saúde, conclui-se que os problemas inerentes à centralização dos dados, especialmente em contextos que demandam maior capacidade analítica e envolvem domínios de negócio densos, são significativamente minimizados. Além disso, a descentralização através dos princípios do data mesh em saúde, eleva o nível de maturidade analítica da instituição, promovendo uma mudança paradigmática na forma como os ativos de dados são percebidos, processados e gerenciados. Não apenas efetuando transformações tecnológicas, mas também mudanças organizacionais, que incluem o letramento em torno dessas soluções a gestores e tomadores de decisão; aumentando a adoção da visão de dados como produto e a implementação de uma responsabilidade federada sobre esses produtos, de forma estruturada.

É importante destacar que esta monografia não se aplica a todos os cenários do setor da saúde, visto que, esse setor é diverso, com instituições de diferentes tamanhos, níveis de maturidade tecnológica e necessidades específicas. Nem rejeita completamente a centralização, reconhecendo sua importância para a garantia de controle e conformidade. Assim, a proposta de data mesh aqui apresentada é mais adequada para contextos onde a descentralização traz benefícios claros sobre os problemas mapeados e apresentados nesta monografia.

### **5.1 Contribuições do Trabalho**

Este trabalho apresenta contribuições para o avanço do debate sobre gestão de dados no setor de saúde, propondo a adoção do paradigma sociotécnico do data mesh como uma alternativa ao modelo centralizado tradicional, uma das principais contribuições está na orientação e fortalecimento de uma cultura orientada a dados dentro das organizações, buscando incentivar o letramento em dados e a criação de um ambiente colaborativo nesse segmento. O estudo busca preparar as unidades organizacionais para atuarem de forma autônoma na produção e gestão de produtos de dados, capacitando os gestores e tomadores de decisões sobre os princípios

que regem esse paradigma, apresentando suas responsabilidades e papéis para uma mudança cultural dentro dos seus respectivos domínios.

Outra contribuição desta monografia é a apresentação de uma estratégia em nível gerencial aplicada a um ambiente centralizado existente, detalhando os passos e os planejamentos estratégicos necessários para alcançar com sucesso a transição para um modelo descentralizado. Essa abordagem oferece uma nova perspectiva para profissionais da área de tecnologia e gestão da informação em saúde, como *chief information office* (CIOs), ao propor uma mudança na forma de enxergar os ativos de dados, tratando-os como produtos e sugerindo a delegação federada de responsabilidades sobre o dado dentro da instituição .

## **5.2 Trabalhos Futuros**

Embora esta monografia tenha proposto uma abordagem estratégica para a análise de viabilidade do data mesh no contexto de instituições de saúde, com um modelo de trabalho simplificado visando a capacitação dos gestores e tomadores de decisão sobre esse paradigma. Ainda existem inúmeras oportunidades para expandir os estudos sobre o tema.

Pesquisas futuras podem focar na elaboração de metodologias mais robustas para viabilizar a transição de um modelo centralizado para um modelo distribuído. Isso inclui a definição de critérios para a escolha de um domínio piloto mais detalhado, a formulação de métricas de sucesso e a identificação de estratégias para minimizar resistências culturais dentro das organizações. Com planos detalhados de como implementá-las e testes em pequenas equipes, avaliando o seu resultado.

Além disso, é relevante investigar casos de uso em larga escala para compreender os limites e as oportunidades de escalabilidade da arquitetura data mesh, especialmente em organizações com um grande número de domínios e alto volume de dados globais, explorando como o data mesh pode coexistir com outras arquiteturas de dados, como Data Warehouses e data lakes e propondo padrões que assegurem a troca eficiente de informações entre esses sistemas.

## REFERÊNCIAS<sup>1</sup>

AMAZON. **Mais vendidos na categoria base de dados**. Amazon. 2024. Disponível em: [https://www.amazon.com.br/gp/bestsellers/books/7842643011/ref=pd\\_zg\\_hrsr\\_books](https://www.amazon.com.br/gp/bestsellers/books/7842643011/ref=pd_zg_hrsr_books) Acesso em: 02 Dez. 2024.

AWS. **O que é uma malha de dados?**. Amazon. 2023. Disponível em: <https://aws.amazon.com/pt/what-is/data-mesh>. Acesso em: 19 set. 2024.

BELLE, A.; THIAGARAJAN, R.; SOROUSHMEHR, S. M. REZA.; NAVIDI, F.; BEARD, D.; NAJARIAN, K. **Big Data Analytics in Healthcare**. Estados Unidos: Hindawi Publishing Corporation, 2015. 16 p.

BRASIL. Lei n.º 13.709, de 14 de agosto de 2018. **Dispõe sobre a proteção de dados pessoais e altera a Lei n.º 12.965**, de 23 de abril de 2014 (Marco Civil da Internet). Diário Oficial da União: seção 1, Brasília, DF, 15 ago. 2018. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/L13709.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm). Acesso em: 25 set. 2024.

DEGHANI, Z. **Data Mesh: Entregando valor em escala e orientado a dados**. 1. ed. São Paulo: Novatec Editora Ltda, 2023

DEGHANI, Z. **How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh**. Martin Fowler. 2019. Disponível em: <https://martinfowler.com/articles/data-monolith-to-mesh.html>. Acesso em: 19 set. 2024.

DPG MEDIA IT. **Data mesh at DPG Media**. Medium, 2021. Disponível em: <https://dpgmedia-engineering.medium.com/data-mesh-at-dpg-media-dfeb612087>. Acesso em 18 nov. 2024.

ESTADOS UNIDOS. **Health Insurance Portability and Accountability Act of 1996**, Pub. L. No. 104-191, 21 de agosto de 1996. Disponível em: <https://www.govinfo.gov/content/pkg/PLAW-104publ191/html/PLAW-104publ191.htm>. Acesso em: 25 set. 2024

GARTNER. **Big Data**. Gartner. 2024. Disponível em: <https://www.gartner.com/en/information-technology/glossary/big-data>. Acesso em: 21 set. 2024.

GENTNER, T.; NEITZEL, T.; SCHULZE, J.; GERSCHNER, F.; THESSLER, A. **Data Lakes in Healthcare: Applications and Benefits from the Perspective of Data Sources and Players**. 27. ed. Grécia: Springer, 2023. 10 p.

GLOBO. **Data Mesh: escalando engenharia de dados | Éfrem de Aguiar Maranhão Filho @ TDC Transformation**. Youtube, 2021. Disponível em: <https://www.youtube.com/watch?v=KZPQQZgFGPI>. Acesso em: 17 nov. 2024

<sup>1</sup> De acordo com a Associação Brasileira de Normas Técnicas (ABNT NBR 6023)

GOOGLE. **Arquitetura e funções em uma malha de dados.** Google Cloud. 2024a. Disponível em: <https://cloud.google.com/architecture/data-mesh>. Acesso em: 28 nov. 2024

GOOGLE. **O que é governança de dados?** Google Cloud. 2024b. Disponível em: <https://cloud.google.com/learn/what-is-data-governance>. Acesso em: 20 nov. 2024

ORACLE. **O que é Big Data?** Oracle. 2024. Disponível em: <https://www.oracle.com/br/big-data/what-is-big-data>. Acesso em: 20 set. 2024.

SINHA, S. **Connected IoT device market update—Summer 2024.** Iot Analytics. 2024. Disponível em: <https://iot-analytics.com/number-connected-iot-devices/> Acesso em: 01 Dez. 2024

SHANKAR, N.; MEYERS, I.; MITCHELL Z.; HASSON R. **Design a data mesh architecture using AWS Lake Formation and AWS Glue.** Amazon. 2021. Disponível em: <https://aws.amazon.com/pt/blogs/big-data/design-a-data-mesh-architecture-using-aws-lake-formation-and-aws-glue/>. Acesso em: 28 Nov. 2024

SKELTON, M.; PAIS, M. **Team Topologies: Organizing Business and Technology Teams for Fast Flow.** 1. ed. Estados Unidos: It Revolution Press, 2019.

<sup>1</sup> De acordo com a Associação Brasileira de Normas Técnicas (ABNT NBR 6023)