

**UNIVERSIDADE DO ESTADO DE SANTA CATARINA - UDESC
CENTRO DE CIÊNCIAS DA ADMINISTRAÇÃO E SOCIOECONÔMICOS - ESAG
MESTRADO PROFISSIONAL EM ADMINISTRAÇÃO**

MARCELO MELLO DO COUTO

**MODELO CONCEITUAL DE GOVERNANÇA DE DADOS: UMA
PROPOSTA PARA O OBSERVATÓRIO**

**FLORIANÓPOLIS
2024**

**UNIVERSIDADE DO ESTADO DE SANTA CATARINA - UDESC
CENTRO DE CIÊNCIAS DA ADMINISTRAÇÃO E SOCIOECONÔMICOS - ESAG
MESTRADO PROFISSIONAL EM ADMINISTRAÇÃO**

**MODELO CONCEITUAL DE GOVERNANÇA DE DADOS: UMA PROPOSTA PARA
O OBSERVATÓRIO**

Dissertação de Mestrado apresentada ao
curso de Mestrado Profissional em
Administração do Centro de Ciências da
Administração e Socioeconômicos
(ESAG), da Universidade do Estado de
Santa Catarina (UDESC), como requisito
final para a obtenção do grau de Mestre em
Administração
Orientador: Júlio da Silva Dias

**FLORIANÓPOLIS
2024**

**Ficha catalográfica elaborada pelo programa de geração automática da
Biblioteca Universitária Udesc,
com os dados fornecidos pelo(a) autor(a)**

Mello do Couto, Marcelo

Modelo conceitual de governança de dados:
Uma proposta para o Observatório / Marcelo Mello
doCouto. -- 2024.

80 p.

Orientador: Júlio da Silva Dias

Dissertação (mestrado) -- Universidade do
Estado de Santa Catarina, Centro de Ciências da
Administração e Socioeconômicas - ESAG,
Programa de Pós-Graduação Profissional em
Administração, Florianópolis, 2024.

1. Governança de dados. 2. Modelo. 3.
Observatórios. I. da Silva Dias, Júlio. II. Universidade
do Estado de Santa Catarina, Centro de Ciências da
Administração e Socioeconômicas - ESAG, Programa
de Pós-Graduação Profissional em Administração. III.
Título.

MARCELO MELLO DO COUTO

**MODELO CONCEITUAL DE GOVERNANÇA DE DADOS: UMA
PROPOSTA PARA O OBSERVATÓRIO**

Dissertação de Mestrado apresentada
ao Curso de Mestrado Profissional em
Administração do Centro de Ciências
da Administração e Socioeconômicas
da Universidade do Estado de Santa
Catarina, como requisito final para
obtenção do título de Mestre em
Administração.

BANCA EXAMINADORA

Orientador: Prof. Dr. Júlio da Silva Dias
Universidade do Estado de Santa Catarina

Prof. Dr. Denilson Sell
Universidade do Estado de Santa Catarina

Profa. Dra. Marli Dias de Souza Pinto
Universidade Federal de Santa Catarina

Florianópolis, 26 de setembro de 2024.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por me guiar e sustentar em toda a minha trajetória até este momento.

À minha família, especialmente aos meus pais, pelo apoio incondicional a todas as minhas escolhas e pelos valores que me foram transmitidos, e que levarei comigo por toda a vida.

Ao meu irmão, minha cunhada e meus sobrinhos, por serem parte essencial da minha vida, sempre me cercando de carinho e amor.

Ao meu orientador, Júlio da Silva Dias, por sua prontidão em me orientar e por compartilhar seu vasto conhecimento, sempre buscando elevar a qualidade deste trabalho.

À minha amiga Dorzeli Salete Trzeciak, por me auxiliar na construção deste trabalho, dedicando seu tempo, inclusive aos finais de semana, para me aconselhar sobre os melhores caminhos a seguir. Além de ser uma amiga parceira em todas as horas, você é uma pessoa extraordinária, uma das mais incríveis que conheço, e sou grato a Deus pela sua amizade.

Aos meus gestores, Alexandre Moraes Ramos e Eliza Coral, por autorizarem a realização deste trabalho no Observatório e por compreenderem a relevância deste tema para a organização.

Aos meus colegas de trabalho, que de alguma forma colaboraram com a realização deste trabalho.

A todas as pessoas que sempre acreditaram em mim, meu sincero agradecimento pelo incentivo constante e por entenderem a importância que este mestrado representa na minha vida.

RESUMO

Atualmente, os dados são vistos como um dos principais ativos de uma organização e o aumento da capacidade de armazenamento e processamento desses dados são os principais motivadores desse movimento. Neste contexto, destacam-se os observatórios que trabalham com dados, principalmente dados abertos governamentais, para subsidiar os estudos e as análises que são oferecidos ao seu público-alvo. É o caso do Observatório, que utiliza esses dados, principalmente relacionados à economia, para fornecer informações para subsidiar o processo decisório dos gestores das indústrias de suas respectivas regiões. No entanto, a transformação de dados em informação é complexa e demanda um conjunto de regras e procedimentos bem definidos para que o dado não perca valor durante esse processo. Neste sentido, o objetivo deste trabalho é propor um modelo conceitual de governança de dados para o Observatório. É uma pesquisa aplicada, de abordagem qualitativa, de caráter exploratório e descritivo. Para a realização dessa pesquisa foi utilizado o método *Design Science Research*, seguindo as etapas: identificação do problema, conscientização sobre o problema, revisão na literatura sobre o tema, proposição e construção do artefato, sua validação pelo gestor e, por fim, a comunicação dos resultados. Como resultado da pesquisa, foi gerado um modelo conceitual contemplando cinco dimensões: qualidade de dados, gestão de metadados, segurança de dados, stewardship e arquitetura de dados, gerenciados por um comitê multidisciplinar que estabelece os princípios norteadores dessas dimensões. Além disso, foram considerados quatro fatores influenciadores a serem monitorados pelo comitê, visando a sustentabilidade do modelo. O modelo desenvolvido foi avaliado pelo gestor do Observatório que o considerou aderente a área, levando em conta as suas necessidades. Espera-se que esse modelo seja um ponto de referência em governança de dados para o observatório, foco desse trabalho e permita que outras organizações que trabalham com dados possam implementá-lo em suas realidades, auxiliando-as a ter dados mais confiáveis, precisos e padronizados, contribuindo para um processo mais eficiente e seguro na sua utilização.

Palavras-chave: Governança de dados; modelo conceitual de governança de dados; observatórios.

ABSTRACT

Data is currently seen as one of the main assets of an organization, and the increase in the capacity to store and process this data is the main driver of this movement. In this context, observatories that work with data, mainly open government data, stand out to support the studies and analyses that are offered to their target audience. This is the case of the Observatory, which uses this data, mainly related to the economy, to provide information to support the decision-making process of managers of industries in their respective regions. However, transforming data into information is complex and requires a set of well-defined rules and procedures so that the data does not lose value during this process. In this sense, the objective of this work is to propose a conceptual model of data governance for the Observatory. This is an applied research, with a qualitative approach, of an exploratory and descriptive nature. The Design Science Research method was used to conduct this research, following the steps: problem identification, awareness of the problem, literature review on the topic, proposal and construction of the artifact, its validation by the manager and, finally, communication of the results. As a result of the research, a conceptual model was generated covering five dimensions: data quality, metadata management, data security, stewardship and data architecture, managed by a multidisciplinary committee that establishes the guiding principles of these dimensions. In addition, four influencing factors were considered to be monitored by the committee, aiming at the sustainability of the model. The developed model was evaluated by the manager of the Observatory, who considered it to be adherent to the area, taking into account its needs. It is expected that this model will be a reference point in data governance for the observatory, the focus of this work, and will allow other organizations that work with data to implement it in their realities, helping them to have more reliable, accurate and standardized data, contributing to a more efficient and safe process in its use.

Keywords: Data governance; conceptual model of data governance; *observatories*.

LISTA DE FIGURAS

Figura 1 - Áreas de conhecimento do <i>framework</i> DAMA.....	15
Figura 2 - Modelo conceitual de governança de dados do Observatório	45

LISTA DE QUADROS

Quadro 1- Frameworks de governança de dados	12
Quadro 2 - Mecanismos dos elementos do modelo	53
Quadro 3 - Implementação do modelo conceitual.....	55

SUMÁRIO

1	INTRODUÇÃO	3
1.1	TEMA E PROBLEMA DA PESQUISA	3
1.2	OBJETIVO GERAL.....	6
1.3	OBJETIVOS ESPECÍFICOS	6
1.4	JUSTIFICATIVA	6
2	FUNDAMENTAÇÃO TEÓRICA	9
2.2	GOVERNANÇA DE DADOS	9
2.2.1	Frameworks de governança de dados.....	11
2.2.1.1	<i>DAMA-DMBOK Framework</i>	14
2.2.1.1.1	Governança de dados	15
2.2.1.1.2	Gerenciamento da arquitetura de dados	15
2.2.1.1.3	Desenvolvimento de dados	17
2.2.1.1.4	Gestão de operações de dados.....	18
2.2.1.1.5	Gerenciamento da segurança de dados.....	20
2.2.1.1.6	Gerenciamento de dados mestres e de referência.....	22
2.2.1.1.7	Gerenciamento da <i>data warehousing</i> e <i>business intelligence</i>	24
2.2.1.1.8	Gerenciamento de documentos e conteúdo	25
2.2.1.1.9	Gerenciamento de metadados	26
2.2.1.1.10	Gerenciamento da qualidade de dados	28
2.3	DADOS ABERTOS E DADOS GOVERNAMENTAIS ABERTOS	30
2.4	OBSERVATÓRIOS	32
3	METODOLOGIA	34
3.1	CARACTERIZAÇÃO DA PESQUISA	34
3.1.1	Identificação do problema	36
3.1.2	Conscientização do problema.....	36
3.1.3	Revisão da literatura	37
3.1.4	Identificação dos artefatos	37
3.1.5	Proposição de artefatos.....	38
3.1.6	Avaliação do artefato	39
3.1.7	Projeto e desenvolvimento do artefato.....	40
3.1.8	Explicitação das aprendizagens	41
3.1.9	Comunicação dos resultados.....	41

4	PROPOSTA DE MODELO CONCEITUAL DE GOVERNANÇA DE DADOS PARA O OBSERVATÓRIO	42
4.1	O PORQUÊ DO MODELO CONCEITUAL DE DADOS.....	42
4.2	PRESSUPOSTOS PARA A CONSTRUÇÃO DO MODELO CONCEITUAL	43
4.3	CONSTRUÇÃO DO MODELO CONCEITUAL	44
4.3.1	Comitê de governança de dados	47
4.3.2	Princípios	47
4.3.3	Qualidade de dados.....	48
4.3.4	Gestão de metadados	49
4.3.5	Stewardship	49
4.3.6	Segurança de dados.....	50
4.3.7	Arquitetura de dados.....	50
4.3.8	Fatores influenciadores	51
4.4	MECANISMOS PARA OPERACIONALIZAR OS ELEMENTOS DO MODELO.....	52
4.5	IMPLEMENTAÇÃO DO MODELO.....	54
4.6	AVALIAÇÃO DO MODELO CONCEITUAL PROPOSTO	58
5	CONSIDERAÇÕES FINAIS	60
	REFERÊNCIAS.....	63
	APÊNDICE A	71

1 INTRODUÇÃO

A presente seção tem como objetivo apresentar o tema, a problemática, os objetivos geral e específicos e a justificativa para o desenvolvimento da pesquisa.

1.1 TEMA E PROBLEMA DA PESQUISA

Os avanços da tecnologia da informação e comunicação (TIC), nas últimas décadas, possibilitaram um crescimento muito grande no volume de dados (Haneem *et al.*, 2019). Com isso, as organizações cada vez mais tem se beneficiado da utilização de dados para tomar decisões. Além disso, o aprimoramento no uso dos dados tem se mostrado útil e necessário não apenas no contexto organizacional, mas também para a sociedade em geral e para os governos, dentre outras esferas. Mas o caminho entre a geração do dado até ele suportar uma decisão é longo e tortuoso. Envolve diversas variáveis até se tornar relevante e com qualidade para auxiliar o tomador de decisão, tais como a extração, a mineração, o tratamento e outros processos que contribuem para que o dado seja entregue com qualidade até o cliente final. Deste modo, a governança de dados (GD) é uma dimensão fundamental de todo este processo, para garantir a qualidade do dado e orientar o seu uso adequado e responsável na organização (Plotkin, 2020).

De acordo com DAMA (2017), a governança de dados é responsável pela orquestração das pessoas e ações necessárias para que os dados estejam aptos às necessidades estratégicas de cada organização e, assim, apoiar a transformação de dados em conhecimento. Com isso, os dados passam a ser mais valorizados e melhor utilizados dentro da organização.

Janssen *et al.* (2020) comentam que as organizações, tanto públicas como privadas, estão cada vez mais adotando estruturas de governança de dados ou de informações, para delimitar as responsabilidades na organização com relação aos ativos de dados, definição de políticas, práticas e procedimentos que contribuam para o melhor gerenciamento dos dados, aumento da qualidade dos dados, proteção dos stakeholders com relação às exigências legais e regulatórias, otimização dos custos envolvidos, dentre outros aspectos com relação a esses ativos.

Tanto do ponto de vista prático, como acadêmico, a responsabilidade com os dados deve ser universal, e precisa estar alinhada com as necessidades das

organizações, visto que os dados são um ativo da governança de dados e estas duas visões estão extremamente alinhadas, conforme enfatiza Otto (2011, p. 47) ao dizer que governança de dados é “uma estrutura que envolve toda a empresa para atribuir direitos e deveres, a fim de poder tratar os dados como um ativo da empresa”.

O conceito de dado como ativo vem crescendo nos últimos anos, já que os dados hoje se encontram em seus mais diversos formatos, sendo gerados em uma velocidade e volume cada vez maior e, igualmente, ganhando mais valor e importância. Além disso, nem todos os dados precisam ser mantidos para sempre em um *data lake*, ou seja, um repositório centralizado de dados, pois podem ser consumidos para uma análise ou estudo específico (Mikalef, 2020).

A longo prazo, a perpetuação de dados com pouco uso pode implicar em um grande espaço obsoleto e com informações não mais utilizadas, o que impacta nos custos relacionados ao armazenamento dos dados, seja em alguma nuvem ou num servidor *on premises* da empresa. Tudo isso é papel da governança de dados e representa a sua importância em uma organização que tem a cultura de dados na sua estratégia, buscando sempre, tomar decisões de modo mais assertivo (Huff; Lee, 2020).

Para apoio ao processo de tomada de decisão, os observatórios surgem como uma ferramenta que pode fornecer subsídios e elementos estratégicos para auxiliar no processo decisório, pois tem, dentre as suas atribuições, ser um centro de informação com atividades de coleta, tratamento e difusão de dados relacionados a uma determinada temática (Reyes; Fernández, 2019). Neste sentido, os observatórios possuem uma relação próxima com dados abertos governamentais, visto que são fontes para análise de cenários e construções de indicadores, que possibilitam entender a capacidade de setores, temáticas ou territórios em períodos determinados, além de prestar serviços de informações com valor agregado para auxiliar a tomada de decisão dos diversos atores envolvidos (universidades, governos, empresas etc.) (De La Vega, 2002).

Atualmente, existem observatórios com foco em diversas temáticas, tais como ciência, tecnologia e inovação, saúde, social, software, prospecção, transporte e logística, dentre outros. E cada observatório tem uma finalidade específica, dependendo do ambiente em que foi criado, da sua natureza e do público-alvo a ser atendido. Essas características afetam e interferem na forma como os observatórios

são estruturados, bem como nos processos, metodologias e procedimentos que são utilizados (Soares; Ferneda; Do Prado 2018).

Este trabalho se desenvolve no contexto de observatórios das indústrias, que tem como missão principal subsidiar o processo decisório dos industriais, por meio da oferta de soluções baseadas em estruturas inteligentes, algoritmos e indicadores, apresentando informações fundamentais ao fortalecimento da indústria em diferentes formatos, tais como: *dashboard*, *data visualization*, relatórios, *report*, tabelas, gráficos, linhas do tempo, *big data*, mapas, dentre outros.

O Observatório é uma plataforma de inteligência com dados, tecnologias, soluções e experiências para suporte às decisões que impulsionam a competitividade e o desenvolvimento sustentável da indústria e seus stakeholders. Possui uma infraestrutura de *big data*, *data analytics* e sistemas de inteligência, abastecidos por cerca de 100 bases dados públicas e privadas do Brasil e do mundo, garantindo a oferta de uma diversidade de informações qualificadas para apoiar o processo decisório de industriais.

Nesse sentido, a presente pesquisa se dará no âmbito de uma organização que trabalha majoritariamente com dados abertos governamentais, o que traz um nível de complexidade maior do que se fosse trabalhado apenas com dados internos. Essas bases possuem peculiaridades distintas entre elas, tanto na periodicidade de atualização, como na estrutura dos seus dados.

Além disso, pelas características da área, o Observatório possui produtos de dados que são ofertados para empresas, por meio de relatórios de listas de empresas, análises econômicas, dados de quantidade de empresas, exportações, previsões de inflação e crescimento econômico etc. Também existem demandas urgentes, que o tempo de entrega é curto, mas que o que é entregue precisa estar de acordo com o esperado. Geralmente, estas demandas são de diretores, vice-presidentes e presidente da instituição. Estas solicitações possuem naturezas bem diversas e exigem um vasto conhecimento sobre as bases de dados abertas disponíveis, além de capacidade para trabalhar com grande volume de dados, o que aumenta a complexidade da sua realização. Tudo isso, exige uma flexibilidade e organização da equipe para atender as demandas com qualidade e no tempo certo.

Outra característica importante do trabalho do Observatório com dados é a sua vulnerabilidade de trabalhar com tantas bases de dados abertos governamentais que mudam sua estrutura com frequência e que geram um desafio para a equipe de fazer

a manutenção do processo que consome esses dados e garantir que ela fique atualizada de forma correta.

Todas essas características citadas anteriormente reforçam a importância de estruturar a governança de dados nas organizações, considerando as melhores práticas disponíveis, mas adaptando-as para o ambiente em que a pesquisa será aplicada. Assim, um modelo conceitual de governança de dados auxilia as organizações a ter uma forma estruturada de representação dos seus processos relacionados a dados, existem diversos modelos e *frameworks* consolidados na literatura como DAMA (2017) e DGI (2021), porém dadas as características citadas em relação a dados do Observatório, faz-se necessário um modelo conceitual, para ser utilizado como referência, personalizado para os desafios específicos da área.

Considerando o exposto, surge a seguinte questão de pesquisa: De que forma e a partir de quais elementos pode ser estruturado um modelo conceitual para a governança de dados do Observatório?

1.2 OBJETIVO GERAL

Propor um modelo conceitual de governança de dados para o Observatório.

1.3 OBJETIVOS ESPECÍFICOS

- a) Identificar técnicas, ferramentas e boas práticas para governança de dados;
- b) mapear o processo atual de governança de dados do Observatório;
- c) identificar os elementos de governança de dados presentes em modelos e *frameworks* para aplicação no Observatório;
- d) estabelecer as diretrizes para implementação do modelo conceitual para o Observatório.

1.4 JUSTIFICATIVA

Os dados devem ser considerados com um ativo para a organização. Mas para que sejam utilizados com eficiência e eficácia, é necessário haver um processo de gestão desses dados, com regras e normas que garantam a sua boa utilização.

De acordo com DAMA (2017) a governança de dados tem se tornado uma estrutura comum em várias organizações que se utilizam de múltiplas tecnologias para compor o *big data*. Além disso, a importância da segurança dos dados cresceu exponencialmente com o aumento dos recursos disponíveis para analisar uma grande quantidade de dados e informações, que fazem parte do cotidiano das organizações.

Uma estrutura de governança e gestão de dados ajuda a organização a ter um controle maior sobre seus ativos de dados. Khatri e Brown (2010) ressaltam que a construção e a delimitação de uma estrutura de governança eficaz é um desafio para as organizações. Atualmente, há muitas ferramentas e tecnologias para trabalhar com dados, mas a sua simples adoção não é suficiente para governar os ativos de dados e de informação (Ladley, 2019). Além disso, práticas e normas excessivas de governança de dados podem prejudicar a eficácia dos processos relacionados a dados.

A presente dissertação considera a governança de dados como um instrumento essencial para organizações que trabalham com dados e que os utilizam não apenas como instrumento para tomada de decisão, mas também como um serviço que pode ser oferecido a outros clientes.

Conforme já mencionado, o Observatório oferece soluções como: prospectivas estratégicas territoriais e setoriais, soluções de inteligência artificial para suporte ao processo decisório, estudos econômicos, análises conjunturais, pesquisa e sondagens industriais, entre outras. Todas essas soluções necessitam de dados confiáveis e íntegros para manter a qualidade dos entregáveis aos seus clientes.

Atualmente, existe um *big data* hospedado em nuvem, que carece de regras, normas, padrões e definições de papéis e responsabilidades para garantir uma boa utilização e organização do dado dentro desse ambiente. Portanto, faz-se necessário um processo de estruturação da governança dos dados do *big data*, e também a criação de um conjunto de elementos que irão guiar a equipe em sua estruturação e no gerenciamento dos dados.

Ademais, o Observatório possui projetos e produtos de dados, que necessitam de manutenção corretiva e evolutiva, além de atualizações mensais, o que acaba ocupando boa parte do tempo dos analistas e exigindo da estrutura de *big data*, dados de qualidade e atualizados.

Nesse cenário, acredita-se que a proposta de modelo para governança de dados possibilitará avaliar os pontos que precisam ser desenvolvidos dentro da GD.

Assim, os benefícios esperados com a proposta de modelo são:

- a) ambiente organizado para utilização e análise de dados;
- b) redução de custos com armazenamento, processamento e gerenciamento de dados;
- c) maior confiabilidade nos dados;
- d) segurança no armazenamento, processamento e fornecimentos dos dados;
- e) maior capacidade de atendimento da equipe a diversas demandas do Observatório.

Cabe destacar que o modelo proposto poderá ser aplicado em outros tipos de observatórios, além daqueles relacionados às federações das indústrias, que utilizam dados abertos governamentais, considerando suas características e diretrizes organizacionais.

Diante do exposto, a relevância desse trabalho se justifica a partir da importância do tema, tanto para o meio empresarial (foco do mestrado) quanto para o meio acadêmico.

Para a academia, espera-se que o modelo para governança de dados gerado traga contribuições, na medida em que dará mais visibilidade ao tema e a possibilidade de desenvolvimento de pesquisas futuras.

Por fim, justifica-se a motivação pessoal para realizar tal pesquisa, pelo fato de contribuir para aumentar o conhecimento do presente autor, resultando em melhor qualificação que, certamente, permitirá melhor atuação profissional.

2 FUNDAMENTAÇÃO TEÓRICA

A revisão de literatura tem como objetivo evidenciar as abordagens e teorias que sustentam a temática escolhida, permitindo formar a base de sustentação para o desenvolvimento do modelo proposto. Desta forma, os principais conceitos desenvolvidos são: governança de dados, modelos de governança de dados, dados abertos governamentais e observatórios.

2.2 GOVERNANÇA DE DADOS

Governança de dados é o exercício de autoridade e controle sobre o gerenciamento dos dados (DAMA, 2017). O seu objetivo é maximizar a percepção dos gestores de que os dados são importantes ativos dentro de uma organização (Ladley, 2019). Porém, atualmente, poucas organizações tratam os seus dados como ativos. Estas organizações geram, processam e utilizam dados das mais variadas formas, mas faltam normas e políticas que regulem a sua gestão.

Recentemente, a governança de dados tem recebido muita atenção relacionada à gestão dos dados e, também referente a outras áreas. Governança de dados é um novo termo que surgiu no início deste século, entretanto, na literatura existem vários termos e abordagens que trabalham com dados e informação, como gestão total da qualidade de dados (Wang, 1998), gestão da qualidade de dados (Wang; Strong, 1996), entre outros (Otto, 2012). Contudo, é importante ressaltar a diferença entre governança e gestão de dados.

Governança se refere a decisões que precisam ser tomadas e a quem toma essas decisões, buscando garantir a eficiência da gestão e o bom uso dos recursos. Já a gestão de dados envolve a implementação das decisões tomadas. A definição de governança indica os responsáveis por suportar as normas e decisões relacionadas ao ativo de dados da organização. No entanto, as dimensões das decisões devem ser identificadas para assegurar os direitos e deveres dos usuários (Alhassan; Sammon; Daly, 2018).

A governança de dados define ações gerenciais de todos os processos que envolvem dados, seja na engenharia, arquitetura ou processamento de dados. Também estabelece controles e procedimentos associados à tomada de decisão baseada em dados.

Alguns autores apresentam governança de dados como um *framework* que auxilia a gestão de dados. O *framework* mais utilizado para gestão de dados é o DAMA-DMBOK, desenvolvido pela *Data Management Association International* (DAMA), que é uma coleção de processos amplamente aceitos como as melhores práticas dentro de setores relacionados a gestão de dados (Bento; Neto; Cortê-Real, 2022).

A DAMA foi fundada em 1980 na cidade de Los Angeles nos Estados Unidos, com o objetivo de ajudar profissionais de dados a aumentar seus conhecimentos e habilidades no gerenciamento de dados, além de disseminar boas práticas, educar e certificar profissionais da área. Ela conceitua a governança de dados como um *framework* que é transversal a vários setores para gerir os dados como um ativo estratégico para a organização (DAMA, 2017).

Abraham, Schneider e Brocke (2019), sugerem que o propósito da governança de dados é buscar melhoria na segurança das informações, redução de custos, alinhamento entre as atividades da empresa e os seus dados e controle da qualidade dos dados. Governança de dados não é uma ferramenta técnica, mas sim um conjunto de normas, boas práticas e políticas para a utilização e distribuição dos dados. A criação de um *big data* sem uma organização e regras bem estabelecidas pode gerar altos custos e perda de informação para a organização.

O objetivo que qualquer sistema de governança de dados deve perseguir é organizar todas as atividades e serviços do *big data* visando criar valor para os dados, a fim de suportar as análises gerenciais. Portanto, o objetivo da governança de dados em um contexto de *big data* deve ser implementar processos que evitem comportamentos inesperados, como vazamento de informações sigilosas, desconfiança nos dados apresentados e ineficiência na busca e recuperação de informações (Kim; Cho, 2017). A infraestrutura de *big data* pode variar em escala, dependendo do modelo de negócio e das estratégias estabelecidas, mas deve ser capaz de fornecer todas as especificidades técnicas para atingir seus objetivos. A auditoria e o controle do sistema devem monitorar os procedimentos e alinhamentos para assegurar que o que foi planejado está, de fato, sendo executado. Se o monitoramento indicar que uma imediata interferência é necessária, essa interferência será feita de acordo com as políticas e procedimentos pré-estabelecidos pela governança (DAMA, 2017).

Uma empresa que deseja ter uma governança em seus dados necessita de um modelo adaptado às suas necessidades, além de um mapeamento de onde estão localizados os dados e uma estratégia para coleta de dados, a partir de diferentes sistemas que geram e armazenam dados dentro da organização (Kim, 2011). A decisão de controle de acessos, responsabilidades e papéis, deverá estar definida e especificada pelo modelo de governança da organização. Este modelo também é necessário para definir uma política de qualidade de dados, distribuir os dados dentro dos setores da companhia e fornecer normas e padrões para todos que consomem os dados dentro da organização.

Entre os componentes da governança de dados, a definição de padrões é fundamental para a qualidade dos dados da organização, porque ajuda a evitar duplicidade, incompletude e desconfiança nas informações fornecidas a partir dos dados. Além disso, como bem colocam Kim e Cho (2017), com um modelo de padrões bem definidos é possível selecionar com mais precisão as tecnologias e ferramentas que serão utilizadas no gerenciamento dos dados.

2.2.1 *Frameworks* de governança de dados

Apesar de a governança de dados ter emergido na literatura acadêmica no início dos anos 80, ainda existem poucos *frameworks* relacionados ao tema (Fu *et al.*, 2011). Esses autores argumentam que uma governança de dados efetiva pode ajudar a esclarecer os objetivos da organização quanto aos dados, aumentar a confiança no uso dos dados por parte dos colaboradores, estabelecer responsabilidades, manter o foco e o escopo, além de definir critérios mensuráveis de sucesso no gerenciamento dos dados.

Para pesquisadores e membros de organizações, a necessidade de um programa de governança de dados aumentou o interesse em desenvolver um *framework* que atendesse as necessidades e trouxesse valor para as organizações. Com isso, alguns *frameworks* conceituais e práticos foram criados nos últimos anos, com elementos específicos de saúde, *cloud computing* e *big data*. Outros, direcionaram seu foco para *data quality management* e *data sharing* (Bento; Neto; Cortê-Real, 2022).

Barata (2015) realizou uma revisão sistemática da literatura em busca de *frameworks* de governança de dados. Segundo esse autor, a maioria dos *frameworks*

encontrados na literatura são adaptados para entender a necessidade específica da pesquisa.

Na literatura pesquisada, foram encontrados sete *frameworks* que podem ser considerados aplicáveis a outros contextos: DAMA DMBOK, Dataflux, Gartner, IBM, *framework* de governança de dados para gestão de ativos, *Framework* de governança de dados para domínio de decisão de dados e, Notre Dame.

No Quadro 1 apresenta-se a descrição de cada um dos sete *frameworks*.

Quadro 1- Frameworks de governança de dados

Frameworks	Descrição
DAMA DMBOK	Proporciona uma visão geral sobre gerenciamento de dados, por meio da definição de padrões, terminologias e melhores práticas, mas sem o detalhe de técnicas e métodos específicos (DAMA, 2017).
Dataflux	O objetivo deste <i>framework</i> é guiar a organização de um programa de governança de dados indisciplinado até um programa integrado e maduro. Possui quatro estágios de maturidade: indisciplinado, reativo, proativo e governado. Este <i>framework</i> destaca a consolidação e a integração da informação (Haider; Haider, 2013; Fisher, 2007).
Gartner	Também conhecido como <i>Enterprise Information Management</i> (EIM), possui foco em análise da maturidade em governança de dados da organização, com seis estágios: inconsciente, consciente, reativo, proativo, gerenciado, proativo e efetivo (Newman; Logan, 2008).
IBM	Possui três dimensões: tipo de <i>big data</i> , disciplinas de governança de dados, indústrias e funções. Ele é bastante utilizado devido a sua adaptabilidade a diferentes contextos de <i>big data</i> e diferentes tipos de organizações (Soares, 2012).
<i>Framework</i> para governança de dados para gestão de ativos	Divide o ciclo de vida dos ativos em sete perspectivas: competitividade, desenho, operações, suporte, <i>stakeholders</i> , aprendizado e eficiência. Incorpora aspectos como qualidade dos dados, integração, padronização, interoperabilidade e gerenciamento de riscos. Possibilita a avaliação do <i>status</i> da organização em relação a governança de dados, além de funcionar como um roteiro para a sustentabilidade e melhoria contínua da governança de dados na organização (Haider; Haider, 2013).
Governança de domínio de decisões de dados	É baseado em cinco domínios de governança de dados que se relacionam entre si: qualidade de dados, acesso aos dados, ciclo de vida dos dados, metadados e princípios de dados. Este <i>framework</i> é adaptado para diferentes níveis de organização, seja ela mais descentralizada ou centralizada, em que, por meio dos domínios de decisão, é possível elaborar perguntas com o objetivo de obter decisões a serem tomadas pelos gestores (Khatri; Brown, 2010).

Notre Dame	Possui cinco pilares: qualidade e consistência, políticas e padrões, segurança e privacidade, <i>compliance</i> , retenção e arquivamento. Estes cinco pilares são suportados por dois importantes domínios da governança de dados: acesso aos dados e tecnologias (Chapple, 2013).
------------	---

Fonte: Elaborado pelo autor (2023)

Estes foram os *frameworks* encontrados até 2015. Para trabalhos a partir de 2015 o autor deste trabalho realizou uma revisão na literatura em três bases de dados: *Web of Science*, *Google Scholar* e *IEEE*. Utilizando as palavras chaves “*framework*” e “*data governance*”. Por meio dessa revisão de literatura, foi possível perceber que existem muitos *frameworks* de governança de dados sendo desenvolvidos especificamente para *big data* (Kim; Cho, 2017; Mashoufi *et al*, 2023; Al-Badi; Tarhini; Khan, 2018), além de *frameworks* específicos para tecnologia e setores da indústria, como *blockchain*, *internet of things*, indústria 4.0, *cloud*, entre outros (Karisma; Tehrani, 2023; Dasgupta; Gill; Hussain, 2019; Zorrila; Yebenes, 2022; Al-Ruithe; Benkhelifa, 2017).

Com isso, o único *framework* encontrado que atende ao critério de Barata (2015) foi o proposto por Abraham, Schneider e Brocke (2019). Os autores fizeram uma revisão de 145 artigos científicos e publicações na web no período de 2001 a 2019. Com isso, seis elementos foram considerados fundamentais para o *framework*: mecanismos de governança, escopo organizacional, escopo de dados, domínio de escopo, antecedentes (contexto) e consequências da governança de dados. O foco deste *framework* é a qualidade de dados, ciclo de vida de dados, segurança, arquitetura, metadados, armazenamento e infraestrutura. Abraham, Schneider e Brocke (2019) também sugerem em seu trabalho um roteiro de aplicação do *framework*. Inicialmente, é necessário identificar os antecedentes (contexto) relacionados a organização e que podem influenciar na sua aplicação, na sequência, pode-se levantar as áreas da organização que será aplicado, quais os dados e quais os objetivos da utilização do *framework*, se ele será terá um enfoque mais na qualidade dos dados, na segurança, armazenamento etc. Por fim, este *framework* traz um elemento novo que são os mecanismos de governança, que não é encontrado em outros *frameworks*, mas que auxilia a compreensão de como ele pode ser aplicado e adaptado ao contexto de cada organização

Além desses *frameworks*, foi encontrado o *Data Governance Institute framework (DGI)*, desenvolvido em 2004 pela instituição de mesmo nome. Ele foi

instituído com o objetivo de classificar, organizar e comunicar atividades complexas do processo decisório envolvendo dados empresariais (Thomas, 2009). Este *framework* possui uma estrutura de dez componentes que estão presentes em qualquer programa de governança de dados, do mais simples ao mais complexo. Os primeiros seis componentes trabalham com as funções e descrevem as regras de engajamento empregados pelos componentes de sete a nove representando as pessoas e a estrutura organizacional, já o décimo componente possui os processos que descrevem os métodos usados para governar os dados. Para Thomas (2009) o intuito do DGI é ajudar uma variedade de partes interessadas em dados de diferentes organizações a se unirem com clareza de pensamento e propósito.

2.2.1.1 *DAMA-DMBOK Framework*

A DAMA é uma organização sem fins lucrativos, composta por profissionais, técnicos e estudiosos, que em 2009 desenvolveu o DAMA-DMBOK (*Data Management Body of Knowledge*) com o intuito de fornecer conceitos e modelos de maturidade de capacidade para padronização de: atividades, processos e boas práticas; funções e responsabilidades; entregáveis e métricas; modelo de maturidade (Barata, 2015).

O DAMA-DMBOK *framework* fornece instrumentos e boas práticas para a gestão de dados, com o objetivo de ter um controle para alavancar o uso dos ativos de dados. Sua missão é atender as necessidades de informação de todos os envolvidos da organização em termos de disponibilidade, segurança e qualidade (Barbieri, 2020). A gestão de dados não é uma responsabilidade apenas do setor de tecnologia da informação da organização, mas também dos clientes internos e externos, envolvendo desde a alta direção, que utiliza dados para decisões estratégicas, até o nível operacional, que em muitos casos coletam e produzem dados.

Na visão de Casaes (2019), o *framework* está estruturado em duas visões: função de gestão de dados e elementos ambientais da gestão de dados. A função de gestão de dados é agrupada por atividades em comuns ou próprias de cada agrupamento, com a necessidade de profissionais ou equipes com funções específicas. Já a função de elementos ambientais da gestão de dados representa as variáveis que influenciam a adoção da governança de dados na organização.

O *framework* DAMA-DMBOK está estruturado por meio de dez áreas de conhecimentos relacionados à gestão de dados, como é possível observar na Figura 1. Na sequência, apresenta-se o seu detalhamento.

Figura 1 - Áreas de conhecimento do *framework* DAMA



Fonte: DAMA-DMBOK (2017)

2.2.1.1.1 Governança de dados

Conjunto de processos, políticas e procedimentos que uma organização estabelece para garantir gerenciamento de dados eficaz, eficiente e seguro. Envolve o cruzamento de diversas disciplinas, com foco em qualidade de dados, passando por avaliação, gerência e melhoria de seu uso, levando em consideração aspectos de segurança e privacidade. Para o DAMA-DMBOK (2017), a governança de dados se divide em duas áreas macro: planejamento e controle da gestão de dados.

2.2.1.1.2 Gerenciamento da arquitetura de dados

O gerenciamento da arquitetura de dados é o processo de definir e implementar uma estrutura para gerenciar dados em uma organização. A arquitetura de dados descreve como os dados são processados, armazenados, distribuídos e consumidos em um sistema de informação (Barbieri, 2020). Com isso, os principais objetivos do gerenciamento da arquitetura de dados, segundo DAMA (2017), são:

- entender as necessidades de informação da organização - a arquitetura de dados se relaciona intrinsecamente com a gestão de dados. O foco é a necessidade de informações, ou seja, algo mais elaborado e focado em negócios. Alguns modelos de dados são gerados a partir do levantamento de informações da organização;
- desenvolver e manter o modelo corporativo de dados - o objetivo é assegurar a singularidade dos dados em toda a organização, utilizando definições semânticas, classificações e agregações de dados, a fim de desenvolver um modelo corporativo de dados íntegro de alta qualidade;
- definir e manter a arquitetura de tecnologia de dados - elencando todas as tecnologias utilizadas na gestão de dados, tais como: ERPs, CRMs, softwares utilizados para o *big data*, ferramentas de modelagem, entre outros;
- definir e manter uma arquitetura de integração de dados - por meio do uso de sistemas que permitam o *data lineage*, ou seja, analisar os dados desde a sua origem, passando por todo ciclo de vida, até chegar na sua inativação;
- definir e manter uma arquitetura de *Data Warehousing* e de *Business Intelligence* - proporcionar estruturas de armazenamento, camadas de transformação e integração, até chegar a camada de consumo de informação, por meio de ferramentas de visualização de dados;
- definir e manter uma taxonomia e padrões de nomes de dados para a empresa - é a definição hierárquica, por meio de classes, subclasses, heranças, composição e agregação de dados, a fim de proporcionar uma consistência entre os elementos e modelos da organização;
- definir e manter uma arquitetura de metadados - consiste na definição de um fluxo estruturado de arquitetura de metadados (dicionários, catálogos, glossários, entre outros) para proporcionar uma estrutura de criação, padronização, utilização e manutenção de metadados, com o objetivo de garantir uma coerência e semântica entre todos os níveis analíticos da organização.

2.2.1.1.3 Desenvolvimento de dados

Esta área comprehende as atividades relacionadas aos dados dentro do ciclo do ciclo de desenvolvimento do sistema, incluindo a modelagem de dados, requisitos de dados e projeto, implantação e manutenção do banco de dados (DAMA-DMBOK, 2017). A estrutura desta etapa consiste em:

- modelagem de dados, análise e projeto de soluções:
 - analisar os requisitos de informações,
 - desenvolver e manter modelos conceituais de dados,
 - desenvolver e manter modelos lógicos de dados,
 - desenvolver e manter modelos físicos de dados,
- detalhar o desenho dos dados:
 - desenhar o banco de dados físicos,
 - desenhar os produtos de informações,
 - desenhar os serviços de acessos a dados,
 - desenhar o serviço de integração dos dados,
- desenho da gestão de qualidade e modelagem dos dados:
 - desenvolver a modelagem dos dados e desenho dos padrões,
 - revisar o modelo de dados e qualidade do banco de dados,
 - gerenciar o versionamento e a integração do modelo de dados,
- implementação dos dados:
 - implementar, desenvolver e testar a mudança no banco de dados,
 - criar e manter ambiente de teste para os dados,
 - migrar e converter os dados,
 - construir e testar produtos de informação,
 - construir e testar serviços de acesso a dados,
 - validar requisitos de informação,
 - preparar para a implementação dos dados.

Com o desenvolvimento destas atividades será possível gerenciar e controlar de maneira eficiente todo o desenvolvimento de produtos e serviços que envolvam os dados da organização.

2.2.1.1.4 Gestão de operações de dados

O objetivo da gestão de operação de dados, segundo o DAMA-DMBOK (2017) é planejar, controlar e apoiar os ativos de dados ao longo de seu ciclo de vida, desde a criação, passando pela aquisição (obtenção) até o arquivamento final (*archiving*) e eliminação (*purge*). A estrutura desta etapa consiste em:

- suporte a banco de dados:
 - implementar e controlar ambiente de banco de dados: significa ter controle sobre os diversos sistemas de gestão de bancos de dados (SGBD's) da organização, suas várias instâncias, tanto de SGBD quanto de tecnologias correlatas, gerência e conhecimento de parâmetros e afinamento de SGBD e correlatos, controle de sua conectividade, ajuste e refinamento de outras camadas correlatas que dialogam com o SGBD e controlam o ambiente de *data storage* usado pelos produtos SGBD's e produtos correlatos,
 - obter dados externos de fontes externas: controle de aquisição de dados obtidos de fontes externas (nos diversos formatos), como vem (licenciados ou contratos), de onde vem (fontes), com qual periodicidade chegam, dados de contrato com fornecedores, SLA com fornecedor, entre outros,
 - planejar para a recuperação de dados (*recovery*): planejamento das atividades de *backup* e *recovery*, com definições de procedimentos, periodicidades, tipos de *backups* (integral, incremental), tipos de mídia, SLA definido para tempos máximo de recuperação, tipos de perda, tipos de recuperação, tipos de arquivos a serem resguardados,
 - realizar *backup* e *recovery* de banco de dados: instanciação do plano definido anteriormente, com criação dos *backups*, *logs* de transações, entre outros,
 - monitorar e ajustar aspectos de performance de banco de dados: realizar acompanhamento proativo (preventivamente) e reativo (após acidentes),
 - planejar a retenção de dados: planejar a forma, tempo e tipo de retenção, transferência para mídias secundárias e de eliminação

- de dados, de acordo com políticas de governança de dados ou aspectos regulatórios,
- arquivar, reter e eliminar dados: instanciação do plano de retenção de dados definido anteriormente,
 - suportar aspectos de banco de dados especializados: entender que existe uma grande variedade de sistemas gerenciadores de banco de dados e de FMS (*File Management Systems*), cada qual disponível para certos tipos de necessidades, sejam dados estruturados, ou não estruturados,
 - gerenciamento de tecnologia de dados:
 - entender os requisitos tecnológicos de dados: quais as limitações das tecnologias de dados existentes, quais os requisitos específicos para a tecnologia, habilidades do “*peopleware*” envolvido, implicações de custos e se há equivalentes no domínio de *softwares* livres,
 - definir arquitetura de dados: definir que tipo de software será necessário em cada camada relacionada com dados: bancos de dados convencionais, banco de dados especiais (XML, georreferenciados, NOSQL, bancos de dados multidimensionais) e que outras camadas são necessárias: integração, ferramentas de modelagem, BI, ETL, camadas para geração de dados, entre outros,
 - avaliar tecnologias de dados: inclui a análise de alternativas tecnológicas de dados. Isso pode ser por um processo de gerência de decisão, no modelo MPS.BR ou DAR (CMMI), envolvendo definição de critérios e pesos para a melhor opção, *benchmarks*, visitas, provas de conceito etc.,
 - instalar e administrar tecnologias de dados: na realidade, é a instanciação dos dois últimos itens anteriormente discutidos;
 - controlar e acompanhar aspectos de liderança de tecnologias de dados: considerar a importância do controle de licenças de uso do SGBD, de ferramentas de BI, de ferramentas de integração, de geradores de relatórios, de cubos, de *mining*, entre outros,

- suportar o uso e as dúvidas (pendências) sobre tecnologia de dados: aqui estão concentradas as ações de apoio, suporte e resolução de incidentes associados às camadas de dados, com processos, por exemplo, do modelo ITIL ou MPV-SV, com detalhamento de níveis de apoio.

Com a realização destas atividades, será possível monitorar e gerenciar todas as operações relacionadas a banco de dados, proporcionando maior confiabilidade e qualidade nas informações obtidas a partir dos SGBD's.

2.2.1.1.5 Gerenciamento da segurança de dados

Segundo o DAMA-DMBOK (2017), o gerenciamento de segurança de dados planeja, desenvolve e executa políticas e procedimentos de segurança, com o objetivo de promover autenticação, autorização, acesso controlado e auditoria de dados e informações. A estrutura desta etapa consiste em:

- entender as necessidades de segurança de dados e os requisitos regulatórios associados: os requisitos regulatórios normalmente estão associados do tipo *SOX*, *Basiléia-II*, *Cerp-Act of Australia*, etc.;
- definir políticas de segurança de dados: são as regras e diretrizes maiores que devem ser seguidas pela organização, em termos de segurança da informação. São normalmente definidas por administradores de segurança juntamente com gestores de dados e auditores de segurança interna ou externa e devem ser aprovadas pelo conselho de governança de dados;
- definir padrões de segurança de dados: compreende: padrões de senha, grupos de usuários, padrões de criptografia, guia para acessos externos, protocolos de transmissão pela internet, requisitos de documentação das informações de segurança, padrões de acesso remoto, procedimentos para relatos de incidentes de segurança, padrões de armazenamento e acesso de dados;
- definir permissões e controle de acesso a dados: significa estabelecer um grau de detalhe sobre as atividades, tanto de planejamento, operação, quanto de controle de gestão de segurança de dados;

- gerenciar usuários, *passwords* e membros de grupos de usuários: analisar os diversos usuários, *passwords*, grupos de usuários, privilégios de usuários e/ou grupos, tendo uma fotografia correta dessas entidades e de suas modificações no contexto de dados;
- gerenciar visões e permissões de acessos a dados: envolve a aplicação dos conceitos de *opt in* e *opt out*, ou seja, as atividades para garantir privacidade e segurança de dados por conceder especificamente permissões, por meio de definições explícitas (*opt in*) ou por se restringir a ações específicas, dentro de um leque amplo de alternativas concedido por *default* (*opt out*). Conceitos de visões (*views*) de bancos de dados também são elementos considerados nesse ponto;
- monitorar autenticação de usuários e comportamento de uso: representa o acompanhamento dos acessos, com observação de comportamentos atípicos ou de riscos, que deverão ser reportados aos envolvidos. Isso alimenta as futuras alterações de planos, projetos e políticas de segurança;
- classificar o grau de confidencialidade das informações: classificar o grau de confidencialidade, definindo hierarquias como: informação para acesso geral (todos podem ver); informações somente para uso interno (somente colaboradores podem acessar as informações que poderão ser mostradas ou discutidas no âmbito externo da empresa, porém não copiadas); informações confidenciais (não devem ser compartilhadas por toda empresa); informações confidenciais restritas (somente aberta para certos colaboradores previamente identificados com o status “devem saber”); informações confidenciais registradas (poucos têm acesso e há a estrita necessidade de assinatura de documento de responsabilidade pelo seu uso ou conhecimento);
- auditar a segurança de dados: objetiva a realização de sessões de auditoria com o propósito de analisar, validar, aconselhar e recomendar políticas, padrões e atividades relacionadas à gerência de segurança de dados.

2.2.1.1.6 Gerenciamento de dados mestres e de referência

Segundo o DAMA-DMBOK (2017), o objetivo do gerenciamento dos dados mestres e de referência é planejar, implementar e controlar atividades para garantir a consistência de dados mestres e de referência. Os dados mestres são os dados fundamentais de uma organização e envolvem clientes, fornecedores, colaboradores, contas, locais, entre outros. Os dados de referência são conjuntos de dados que descrevem códigos presentes em tabelas.

Os dados mestres e de referência servem como *input* para os dados transacionais. Os dados de referência podem ser o status do pedido, CEP padrão do fornecedor, entre outros. A sua estrutura é composta por:

- entender a necessidade de integração dos dados mestres e de referência: significa ter os devidos controles para compatibilizar os dados que estão replicados e usados em diferentes sistemas/aplicativos.
- identificar as fontes e contribuinte de dados mestres e de referência: depois de ter entendido, é importante traçar a linhagem do dado, para identificar suas fontes originais e temporárias (*big data*, FMS, processo, área organizacional, pessoas, papéis/indivíduos envolvidos);
- definir e manter a arquitetura de integração de dados: já mencionada anteriormente na função gerência da arquitetura de dados como “definir e manter uma arquitetura de integração de dados”, a arquitetura de solução MDM passa por topologias parecidas com as de bancos de dados distribuídos e/ou replicados e busca evitar o aparecimento de “silos” de dados ou arquivos isolados e personalizados para atender aplicações específicas;
- implementar soluções de gerência de dados mestres e de referência: passa por definição de soluções que permitam o uso compartilhado do dado mestre ou de referência, na sua forma “golden record” por variadas aplicações *Online Transaction Processing* (OLTP) ou de *Business Intelligence* (BI), mantendo a sua integridade. As topologias devem permitir leituras diretas do DM (dados mestres ou de referência) ou o seu uso em sistemas por meio de replicações controladas (síncronas ou assíncronas);

- definir e manter regras de “*match*” para os dados replicados: trabalhar padrões para que se possa identificar duas ocorrências como sendo do mesmo objeto. Tal atividade passa por técnicas de identificação de elementos duplicados, por meio de regras de inferência de similaridade, por técnicas de eliminação de duplicação de registros iguais e por técnicas de consolidação que podem ser o *match merge* (consistem em técnicas de batimento das várias ocorrências multiplicadas) e *match link* (técnicas que definem um registro, com apontadores para outros, sem consolidação física em um único elemento);
- definir os “*golden records*”: “*golden records*” significa o conceito de fonte única, íntegra e confiável, que procura garantir a verdade sobre os dados. Por exemplo, uma fonte de dados para exportações de empresas brasileiras;
- definir e manter hierarquias e afiliações: as hierarquias e afiliações complementam as informações de relacionamentos dos dados mestres, mostrando, por exemplo, dois registros mestres de clientes, relacionados como todo-parte (um cliente é parte de um outro cliente, ou seja, faz parte do mesmo grupo, ou é afiliada, ou seja, tem um relacionamento com a outra empresa). Também, há o relacionamento conhecido como “É um tipo de”. As duas classificações de dados são muito aplicadas no conceito de objetos (todo-parte ou composição e agregação), adotados em modelos de classes e objetos;
- planejar e implementar integrações das novas fontes de dados: nesse ponto o DAMA-DMBOK (2017) se preocupa com a chegada de novas fontes de dados que deverão ser integradas ao ambiente já existente. Isso envolve: analisar as requisições feitas de integração, a complexidade e custo de sua integração e avaliar a quantidade de dados entrantes;
- replicar e distribuir dados mestres e de referência: versa sobre a arquitetura definida para a solução de MDM implementada. Poderá ser via banco de dados distribuídos, ou através de replicações;
- gerenciar alterações de dados mestres e de referência: esse ponto, de crucial importância, deverá ser considerado com cuidado, pois os dados do ambiente MDM estarão compartilhados e as suas alterações

implicam controles mais rigorosos, a fim de evitar impactos e rupturas em seu funcionamento. Passa por: criar e receber pedidos de alteração, identificar áreas interessadas; avaliar impactos em função da alteração, aceitar ou rejeitar a solicitação ou transferir a decisão para o âmbito da governança de dados (GD), comunicar às partes interessadas acerca das mudanças.

Resumindo, os dados mestres são os dados principais de uma organização, representando os objetos negócios de uma organização (Barbieri, 2020).

2.2.1.1.7 Gerenciamento da *data warehousing* e *business intelligence*

Para Mosley *et al.* (2009), o gerenciamento de DW e BI consiste em planejar, implementar e controlar processos para prover dados de suporte à tomada de decisões e apoio aos colaboradores envolvidos na produção de relatórios de conhecimento, consultas e análises. A estrutura desta etapa consiste em:

- entender as necessidades de informações analíticas (BI): os requisitos de BI são revestidos de certa diferença quando comparados com os requisitos tradicionais de sistemas transacionais. O analista de requisitos vem de uma camada mais tática e estratégica de uma organização. Isso demanda uma observação mais apurada sobre as necessidades e problemas (requisitos de negócios e de usuários), além de técnicas mais efetivas de comprometimento das partes, como protótipos, provas de conceito, entre outros. O levantamento das necessidades de negócio sugere a captura de metadados (significado dos dados, dos processamentos, de indicadores, etc.). Nesse momento, é importante a observação comparativa dos dados demandados em relação aos dados existentes;
- definir e manter a arquitetura de DW e BI: passa pelas diferentes alternativas de escolas: visão de EDW (Bill Innomnn) e de DMarts evolutivos e integrados (Ralph Kimball), com todos os elementos que formam uma arquitetura de DW e BI: sistemas transacionais fomentadores dos dados, camada de ETL, camada de armazenamento

com *DataWarehouse*, *DataMarts*, ODS, camada de ferramentas para produção de informações, camada de *profiling* e *cleasing* dos dados, etc.

- implementar os DW e *Data Marts*: nesta atividade o DMBOK foca na materialização gradativa de DW e BI, em projetos evolutivos e integrados;
- implementar as ferramentas de BI e de interface: passa pelos tipos de ferramentas necessários para alcançar os objetivos. São ferramentas de *dashboards*, ferramentas de visualização de dados, ferramentas de relatórios, OLAPs (de cubos), de análise preditiva/*mining*, entre outros, que formarão o arsenal de aplicativos para atender as necessidades de informação da empresa;
- processar os dados para o ambiente de BI: relaciona-se com as atividades de ETL (extração, transformação e carga) de dados, atividades de *cleansing* e integração de dados, considerando áreas intermediárias, como *staging*, dados intermediários como ODS, etc.;
- monitorar e ajustar os processos de DW: passa pelos processos de monitoração de performance de banco de dados que compõem a estrutura do DW, processos de *backup/recovery*, processos de *archiving*, etc.;
- monitorar e ajustar as atividades e aspectos de performance de BI: passa por atividades de monitoração de tempos de respostas de aplicativos de BI, número de usuários de BI por unidade de tempo, entre outros.

2.2.1.1.8 Gerenciamento de documentos e conteúdo

Para o DAMA-DMBOK (2017), o objetivo desta etapa é implementar, planejar e controlar atividades para armazenar, proteger e acessar dados encontrados em arquivos eletrônicos e registros físicos (texto, gráficos, imagens, áudio e vídeo), ou seja, o foco é em dados não estruturados, não armazenados em sistemas relacionais. Há duas grandes subfunções: gerência de documentos e de registros e gerência de conteúdo. Esta etapa é dividida em dois grupos: gestão de documentos e conteúdo e gerência de documentos e registros.

Na gestão de documentos e conteúdo, a gerência está relacionada à gerência da informação e gerência de conhecimentos e trata fundamentalmente de aspectos de entendimento e classificação de conteúdos de documentos, via aplicação de conceitos de taxonomia. São normalmente estruturadas via uma visão ontológica (conhecimento de ser ou entes), com taxonomias da seguinte forma: achatada (com os elementos listados em sequência, sem uma estruturação entre eles), hierárquica (com uma organização dos elementos apresentando certa forma de subordinação), na forma de *facets* ou estrelas (com os elementos dispostos numa forma radial, dispostos em torno de um ponto central, como comumente encontrado nos mapas mentais) e de rede (misturando a hierarquia com *facets*).

Já o gerenciamento de documentos e registros se fundamenta em dois pilares: a gerência do documento em si e o seu conteúdo. O primeiro se preocupa com o documento como se fosse um objeto e o outro com a sua estruturação e conteúdo.

2.2.1.1.9 Gerenciamento de metadados

Na concepção do DAMA-DMBOK (2017, p. 217), o gerenciamento de metadados consiste “em planejar, implementar e controlar atividades que viabilizem um fácil acesso aos metadados integrados e de qualidade”. A estrutura desta etapa consiste em:

- entender os requisitos de metadados: os metadados estão para os dados assim como os dados estão para as “coisas”/entidades colocadas sobre os sistemas de processos de um sistema computadorizado. Por exemplo, os objetos, os eventos, as transações e os relacionamentos são as “coisas” controladas no computador, por meio de sistemas. Assim, os dados definem esses objetos, da mesma forma que os metadados definem os dados. Assim, a gerência de metadados é um processo que controla a sua criação (quando se define, se entende e se documenta aquilo que está sendo objeto do processo), o seu armazenamento (se estrutura, se carrega e se cuida para que seja acessado com facilidade e rapidez), a sua integração (quando dois ou mais metadados sobre o mesmo objeto, diferentemente definidos em tempos distintos, por unidades organizacionais distintas, não estão consistentes) e o seu controle (quando se procura mantê-los com

qualidade e sobre os quais se define métricas, no sentido de que não se controla aquilo que não se mede). Os metadados aumentam o valor da informação estratégica lhe dando expressão, detalhes, conhecimento. Isso reduz o custo de aprendizado, pois as informações sobre os dados estarão mais claras, assim como o tempo gasto na busca pelo entendimento de certos objetos, regras e fórmulas;

- arquitetura de metadados: como a arquitetura de dados, a de metadados pode ser centralizada ou descentralizada, dependendo de como os repositórios de metadados estão dispostos. A centralizada dispõe das vantagens de um controle mais rigoroso e de menor conflito, visando a criação de uma estrutura única e consolidadora. Por outro lado, tem como desvantagem, o esforço de se concentrar os metadados oriundos de diversas origens em um único repositório. A descentralizada tem a vantagem de economizar nos gastos de integração, não havendo persistência centralizada, porém, com um custo de busca integrada, em vários depósitos, para se resolver as consultas solicitadas. Uma arquitetura mista envolve a parte da descentralização, com as buscas dinâmicas e outra parte de definição centralizada no catálogo único;
- desenvolver e manter padrões de metadados: os metadados, basicamente, são de dois tipos: negócios e técnicos. Os metadados de negócios têm o objetivo de documentar os elementos de negócios, centrando num patamar mais conceitual. Envolvem definições e processos de negócios, sistemas, aplicações e aplicativos, regras de negócios, formas de cálculos, algoritmos, linhagem de dados, modelos conceituais e lógicos de dados, aspectos de qualidade de dados e de conceitos de gestores de metadados e das unidades organizacionais responsáveis por eles;
- outros pontos do processo: o desafio de criar e manter metadados é muito grande. Daí decorre a ainda, baixa incidência de implementação nas empresas. Normalmente se tem modelos isolados oriundos das ferramentas adquiridas, sendo a sua integração um dos grandes desafios. A instanciação dessa gerência se dará pelo gerenciamento adequado dos diversos repositórios, que possam produzir, distribuir e

entregar os metadados na forma de consultas, relatórios e análise, no momento exigido e com a devida consistência.

2.2.1.1.10 Gerenciamento da qualidade de dados

Esta etapa é uma das mais centrais em uma governança de dados. A sua estrutura consiste em:

- desenvolver e promover a consciência para qualidade de dados: é preciso difundir a importância dos conceitos, seja por mecanismos diretos ou indiretos. Os diretos seriam por meio de palestras, consultorias ou assemelhados. Os meios indiretos são por meio de exemplos acerca de efeitos negativos da qualidade de dados nas empresas. É importante elucidar que o problema não é exclusivamente de domínio da TI, mas principalmente um problema da esfera de negócios. A qualidade de dados deve ser um dos elementos fundamentais do arco de governança de dados, que define política, padrões, procedimentos, papéis, programas e projetos dentre outros itens do seu escopo, visando tratar e preservar o “ativo” dado. A realização de um trabalho inicial de *profiling* dos dados mais importantes da empresa, mostrando os resultados preocupantes com relação à qualidade de dados é algo a ser considerado pela organização, pois serve de início para todo o processo de convencimento material sobre o problema dos dados com baixa qualidade;
- definir requisitos de qualidade de dados: os requisitos de qualidade de dados são definidos de acordo com a necessidade da organização. Deve-se levar em consideração os processos críticos da organização, suas regras de negócio, seus dados consumidos e produzidos e o impacto da qualidade de dados na sua execução, tanto como *input* como *output*. De acordo com o DAMA-DMBOK (2017), esses requisitos são: precisão, completude, consistência, atualidade, precisão numérica, disponibilidade e unicidade;
- estabelecer processos de “*profiling*”, análise e avaliação de qualidade de dados: refere-se a necessária identificação do estágio inicial dos dados em diferentes áreas ou domínios das organizações. É recomendado que

seja uma das primeiras ações para se estabelecer as “*baselines*” dos processos de melhoria de qualidade de dados das organizações. Permite criar as primeiras métricas e definição dos objetivos a serem alcançados a partir delas;

- definir métricas para a qualidade de dados: as métricas deverão ser definidas para a avaliação do estado atual e da evolução dos tratamentos de qualidade de dados. As métricas, como todas as medidas definidas em processos de qualidade de dados deverão: ser atreladas a objetivos bem definidos, responder as questões associadas a esses objetivos, serem medidas com clareza, definir regras de negócios para qualidade de dados. As métricas devem estar intrinsecamente ligadas às regras de negócios da organização, para evitar possíveis quebras de conformidade no seu uso;
- testar e validar os requisitos de qualidade de dados: nesse ponto específico, o DAMA-DMBOK (2017) sugere a verificação constante e recorrente, a fim de que os dados sejam sempre nos seus domínios de qualidade;
- definir e avaliar níveis de serviço de qualidade: neste item, são avaliados os níveis de serviço (SLA) de qualidade de dados, que deverá ser garantido por medições e verificações constantes. Os níveis de SLA são o compromisso firmado sobre qualidade da área gestora de dados com seus usuários. Os itens subsequentes, relativos a medir e monitorar continuamente, gerenciar as pendências e corrigir os defeitos são consequências diretas desse compromisso de nível de serviços;
- projetar e implementar procedimentos operacionais de gerência de qualidade de dados e monitorar os procedimentos operacionais e a performance de gerência de qualidade de dados: é a materialização do processo de garantia de qualidade de dados. Todo processo deverá ser constantemente avaliado, a fim de se obter melhorias nos seus procedimentos, políticas e resultados.

2.3 DADOS ABERTOS E DADOS GOVERNAMENTAIS ABERTOS

Para a *Open Knowledge Foundation* (2023), o dado aberto é conceituado como um dado que qualquer pessoa pode, livremente, acessar, utilizar, reutilizar, modificar e compartilhar. No contexto de gestão pública, o termo “dados abertos” foi mencionado pela primeira vez em um programa de cooperação internacional para operar satélites estadunidenses (POSSAMAI, 2016). Para tal, o programa requereu de entidades governamentais, políticas de “dados abertos”, tais como as empregadas pela *National Aeronautics and Space Administration* (NASA) e outras agências dos EUA, a fim de compartilhar dados e informações de pesquisa.

A *Open Knowledge Foundation* (2023) estabelece três normas fundamentais para considerar dados abertos: a disponibilidade de acesso, o reuso e distribuição e a participação universal. Essas três normas podem ser resumidas da seguinte forma:

- a) se o dado não pode ser encontrado ou indexado na Web, ele não existe;
- b) se o dado não está disponível num formato aberto e legível por máquina, ele não pode ser reutilizado;
- c) se dispositivos legais não permitem que ele seja compartilhado, ele não é útil.

Dados abertos, são dados disponíveis ao público, dos mais diferentes formatos e fontes e fornecidos sob licença aberta que permita sua livre ingestão, análise ou inferência, com a condição estabelecida de referenciar a fonte (Santos, 2020).

No Brasil, a publicação de informações pela Administração Pública Federal brasileira é um dever dos órgãos públicos, pelo princípio da publicidade estipulado na constituição de 1988, bem como o direito do cidadão em receber informações públicas. Com o advento da Lei de Acesso à Informação (LAI), Lei 12.527 de 2011, em 2019, foi lançado o Portal Brasileiro de Dados Abertos com o objetivo de ser o ponto central para a busca, o acesso e o uso dos dados abertos governamentais, além de promover a interlocução de atores da sociedade e o governo para pensar na melhor utilização dos dados em prol de uma sociedade melhor (Portal Brasileiro De Dados Abertos, 2023).

No contexto de dados abertos no Brasil, foi criada a Infraestrutura Nacional de Dados Abertos (INDA), que é um conjunto de padrões, tecnologias, procedimentos e mecanismos de controle necessários para atender controlar o compartilhamento de

dados. Ela foi instituída em 2016, a partir do mesmo decreto que estabeleceu a Política de Dados Abertos do Poder Executivo Federal.

No entanto, manter os dados disponíveis apenas em portais eletrônicos não é suficiente para contribuir com o aumento da transparência dos serviços públicos. É preciso ter uma auditoria constante sobre o que é disponibilizado e mais do que tudo, o usuário precisa ter conhecimento sobre como extrair e manipular esses dados para utilizá-los em benefício próprio e, por conseguinte, da sociedade. A fim de buscar uma alternativa para esse e outros assuntos, um grupo de 30 especialistas representantes de diversas empresas de tecnologia, estabeleceu oito princípios para os dados governamentais abertos (Attard, 2015). Eles devem ser:

- a) completos: todo dado público deve se manter disponível. Dado público é dado, desde que não seja limitado a validações de privacidade, segurança e controle de acesso;
- b) primários: dado é coletado na fonte, com maior granularidade possível, não agregado ou modificado de outra forma;
- c) atualizados: dado precisa ser disponibilizado da forma mais rápida possível para preservar o seu valor;
- d) acessíveis: dado é disponível para uma maior quantidade de usuários, com os mais variados propósitos;
- e) compreensíveis por máquinas: o dado tem uma estrutura razoável que permita automatizar seu processo;
- f) não discriminatórios: dado disponível para todos, sem nenhum pré-requisito para acesso;
- g) não proprietário: dado disponível em um formato que nenhuma entidade tenha controle exclusivo;
- h) livres de licença: os dados públicos não devem estar submetidos a *copyrights*, patentes e marcas registradas.

Todo dado público tem potencial para ser aberto. Entretanto, o dado público não deve estar sujeito a limitações de privacidade, segurança, controle de acesso ou outros privilégios regulados por estatutos. Com isso, eles não compreendem dados pessoais, sob a custódia estatal ou dados sigilosos de organizações privadas. Além disso, alguns dados governamentais, devido a sua sensibilidade e por questões

estratégicas de estado, demandam reserva e segredo, desde que devidamente regulamentadas pelo estado (Possamai, 2016).

Na sequência aborda-se o tema observatórios, por tratar-se do foco para o qual é desenvolvido o modelo de governança de dados.

2.4 OBSERVATÓRIOS

Apesar do surgimento dos observatórios estarem vinculados a observatórios astronômicos e da existência de observatórios na década de 70, os observatórios com objetivos de abordar temas socioeconômicos surgiram na década de 80, na Europa (Ortega; Del Valle, 2010). Eles ainda destacam que o surgimento dos observatórios está vinculado às condições históricas da invenção e popularização dos indicadores e estatísticas associados a uma política.

O surgimento da internet também foi um grande contribuidor para a expansão dos observatórios (Martins, 2007). Com a internet foi possível ampliar exponencialmente as fontes e redes abertas de informação, o que não era possível no período pré-internet em que os observatórios utilizavam sistema unidirecional, por meio de circuitos fechados que não propiciavam interação externa. Com isso, a informação ou o conhecimento chegava atrasado para o público (Martins, 2007).

O conceito de observatório é diverso. As suas definições, características e objetivos conceituais variam sobre o que se propõe a fazer e como pode se caracterizar uma ferramenta como observatório (Barbosa, 2020). Ortega e Del Valle (2010) apontam que não existe uma definição única e que isso dificulta a sua identificação e limita a análise do fenômeno. Além disso, a diversidade organizacional e operacional que um observatório permite aumenta a dificuldade de se ter uma definição única. Assim, seu conceito ao longo do tempo acaba sendo definido conforme o viés do autor e da prática em que o observatório está inserido (Aspinall Jacobson; Castillo-Salgado, 2016; López *et al.*, 2012).

O significado de observação segundo Mora (2001) é vigiar, o que se faz atentamente e com vistas a certos fins. Observatório se caracteriza como lugar ou posição que serve para fazer observações; edifício ou instalação dotados de pessoas ou instrumentos apropriados, dedicado a realizar observações (Espasa, 1995). Com a combinação das percepções dos dois autores, é possível afirmar que observatório é um local com condições técnicas e tecnológicas que permitem visualizar

especificamente e atentamente um fenômeno com um determinado fim (Barbosa, 2020).

Segundo Silva *et al.* (2013), os observatórios podem ser caracterizados como um sistema de informação que tem como intuito a visualização de informações e dados de determinada área, além de propor certas análises baseadas em indicadores importantes sobre uma determinada temática.

Gusmão (2005) e Beuttenmuller (2007) argumentam que algumas ferramentas executam as mesmas tarefas que um observatório, porém, por alguma razão não são definidas como tal.

Os observatórios podem ser classificados em territoriais, que estudam questões referentes a um espaço geográfico, seja cidade, estado ou região, e temáticos, que possuem um foco em aspectos sociais, econômicos ou políticos. Também existem os observatórios que se caracterizam pelos dois tipos, e isso acontece pelo fato da questão geográfica fazer parte de um determinado interesse da sociedade (Santoro; Xavier, 2009).

O observatório tem como um dos seus papéis fundamentais o apoio ao processo decisório, fundamentando a decisão de gestores com dados e informações tratadas e qualificadas, essa, aliás, é uma das atividades dos observatórios: coletar, compilar e tratar informações para subsidiar a tomada de decisão (Gusmão, 2007; Phélan, 2007; Batista *et al.*, 2016).

Salienta-se que para executar as atividades citadas, o observatório deve contar com equipe qualificada, infraestrutura física e tecnológica, além de um banco robusto de dados estruturados e não estruturados, que forneçam subsídios e informações relevantes para apoiar o público-alvo em suas decisões (Silva *et al.*, 2013). Dessa forma, a governança de dados tem um papel fundamental no gerenciamento desses elementos, buscando a harmonia entre eles, a fim de que os objetivos estratégicos dos observatórios sejam atendidos.

3 METODOLOGIA

A presente seção apresenta os procedimentos metodológicos utilizados para o desenvolvimento da pesquisa.

3.1 CARACTERIZAÇÃO DA PESQUISA

O levantamento de dados para o desenvolvimento de uma pesquisa envolve uma escolha adequada de métodos e procedimentos para o contexto analisado. Carvalho (2013) ressalta que é essencial estabelecer uma metodologia para auxiliar na coleta, análise e interpretação dos dados. Neste sentido, a caracterização da pesquisa pode ser classificada quanto a sua natureza, abordagem, objetivos e procedimentos metodológicos.

Em relação a sua **natureza**, trata-se de uma pesquisa aplicada, tendo como característica principal contribuir para fins práticos, aplicar ou utilizar, mais ou menos imediatamente, os resultados na solução de problemas que ocorrem na realidade (Marconi; Lakatos, 2007), no caso, o desenvolvimento de um modelo.

Grove, Burns e Gray (2012) definem um modelo como sendo exemplos de grandes teorias, com construtos relacionados. Um modelo explica amplamente o fenômeno de interesse. Ao contrário do *framework* que promove a geração de teorias relacionadas ao tema e propõe ferramentas e boas práticas para solução de problemas, o modelo tem um objetivo mais simples, que é de representar uma realidade e um fenômeno por meio de figuras ou diagramas (Casaes, 2019). Modelo é um conjunto de representações ou regras representando relações entre teorias e constructos. Ele pode ser visto apenas como uma descrição, que é uma representação de como as coisas são (March; Smith, 1995).

Para esse trabalho, a preocupação maior foi com a utilidade do modelo e não apenas com sua aderência em representar a verdade. Com isso, embora um modelo possa ser impreciso em relação aos detalhes da realidade, ele deve capturar a estrutura geral da realidade, buscando assegurar a sua utilidade (March; Smith, 1995).

Quanto a abordagem da pesquisa, pode ser considerada como **qualitativa**, pois se preocupa com o nível de realidade que não pode ser quantificado, ou seja, trabalha com universo de significados, motivações, crenças, valores e atitudes (Minayo, 2014).

Em relação aos seus objetivos, trata-se de pesquisa **descritiva** pois objetiva analisar as características, fatores ou variáveis que se relacionam com determinado fenômeno. Em relação às etapas de análise do modelo, a pesquisa pode ser considerada como **exploratória**, visto que objetiva desenvolver e esclarecer conceitos e ideias acerca do tema da pesquisa, além de exigir discussão com especialistas e revisão da literatura para elucidação de fenômenos (Perovano, 2014).

Considerando os objetivos gerais e específicos da pesquisa e as informações apresentadas no referencial teórico, foram explorados diversos *Frameworks* e modelos disponíveis na literatura, por meio da metodologia *Design Science Research* (DSR), proposta por Dresch (2013).

Cabe salientar que há diversas adaptações da metodologia DSR. Este estudo utiliza uma adaptação da metodologia proposta por Dresch (2013), que divide a pesquisa em 9 etapas. Essa escolha tem como objetivo a produção do artefato: modelo para governança de dados do Observatório.

A DSR é a metodologia que operacionaliza e fundamenta a condução da pesquisa quando o objetivo a ser alcançado é um artefato, ou ainda, uma prescrição. Ela tem como objetivo, estudar, pesquisar e investigar o artificial, tanto do ponto de vista acadêmico quanto organizacional (Bayazit, 2004).

O artefato é um elemento essencial para o DSR e pode ser classificado em: constructos (definido como um vocabulário de um domínio, constitui uma conceituação utilizada para descrever problemas dentro do domínio e descrever as soluções), modelos (conjunto de proposições ou declarações que expressam relacionamentos entre constructos, podem ser vistos como descrições ou representações de como as coisas são), métodos (conjunto de etapas (algoritmo ou orientação) usadas para executar uma tarefa, é a concretização de um artefato em seu ambiente, operacionaliza constructos, modelos e métodos) e instanciações (informam como implementar determinado artefato e seus possíveis resultados) (March; Smith, 1995).

Para Çağdaş e Stubkjær (2011), a DSR é um processo rígido para desenvolver artefatos, validá-los e comunicá-los. Nesse sentido, as soluções geradas devem ser passíveis de generalização para uma determinada classe de problemas, permitindo que outros pesquisadores e profissionais em situações diversas possam utilizar o conhecimento gerado (Van Aken, 2005). As classes de problemas podem ser definidas como a organização de um conjunto de problemas, teóricos ou práticos, que

contenham artefatos avaliados e úteis para a organização. Elas representam uma característica importante na DSR, pois permitem que os artefatos e suas soluções não sejam algo pontual relacionado a certo problema em um contexto específico (Lacerda *et al.*, 2013).

Para este trabalho, foram utilizadas nove etapas da *DSR* a partir da adaptação do trabalho de Dresch (2013), as quais são apresentadas na sequência.

3.1.1 Identificação do problema

Esta etapa compreende o entendimento do problema de pesquisa, assim como sua relevância e importância (Dresch, 2013). O problema iniciou pela curiosidade do autor da pesquisa em explorar modelos e métodos relacionados a governança de dados. Além disso, foram buscadas aplicações desses modelos para dados abertos governamentais, pois possuem particularidades diferentes de dados internos de organizações, isso geram muita vulnerabilidade para a qualidade dos dados do Observatório, pois são diversas as bases de dados utilizadas que sofrem mudanças em sua estrutura com frequência, exigindo uma flexibilidade dos analistas da área para adaptar o processo a essas mudanças. Também, acredita-se que se faz necessário um modelo de governança de dados para o Observatório, devido a incipiente adoção de um *big data* em nuvem, bem como o crescimento da equipe de analistas que consomem e utilizam os recursos e a infraestrutura para a análise de dados. Além disso, o Observatório recebe demandas institucionais por dados e análises dos diversos setores das Federações. Essas demandas possuem um caráter multidisciplinar, que vai de temas como saúde e educação, até acidentes de trabalho, carecendo de uma dinamicidade da equipe para atendê-la sem tempo e com qualidade. Para isso, é fundamental um ambiente bem-organizado e com padrões e normas estabelecidos para que a entrega atenda às necessidades dos demandantes.

3.1.2 Conscientização do problema

Para Dresch (2013), nesta etapa é necessário o esforço do pesquisador em ampliar o escopo do problema. Para a conscientização do problema, a busca de informações sobre o tema é feita por meio de uma revisão da literatura, seja da literatura acadêmica, seja da literatura cinza. Também, os líderes, bem como os

analistas, sabem da importância de se ter uma governança de dados bem estabelecida no Observatório, visto que a complexidade das demandas e o nível de qualidade que é exigido do Observatório é alto.

Em um contexto mais amplo, é notório um crescimento da adoção de governança de dados pelas organizações, o Observatório recebe muitas visitas, e em conversas é possível observar que a disponibilidade de dados para consumo aumentou muito junto com as ferramentas para trabalhar com eles. Por outro lado, esse crescimento não vem acompanhado de regras e responsabilidades bem definidas, ocasionando em um *data swamp* (pântano de dados), ou seja, um ambiente de dados com baixa eficiência e que agrega pouco valor ao negócio.

3.1.3 Revisão da literatura

A terceira etapa é fundamental para a busca de conhecimentos existentes e aplicações semelhantes que foram realizadas anteriormente.

Este trabalho tem como foco governança de dados, *frameworks* e modelos de governança de dados, dados abertos governamentais e observatórios. Com isso, foram realizadas buscas em bases de dados tais como *Web of Science*, *Google Scholar* e *Science Direct* a fim de encontrar pesquisas semelhantes ao objetivo deste estudo. As palavras-chave pesquisadas foram “governança de dados”, “*data governance*”, “*frameworks*”, “modelos”, “*models*”, “*observatories*”, “*open data government*” e “dados aberto governamentais”, todos estas palavras-chave foram buscadas de forma isolada, depois elas foram buscadas de forma conjunta com operadores booleanos para buscas trabalhos mais aderentes com o tema desta pesquisa.

3.1.4 Identificação dos artefatos

Para a identificação de artefatos é necessário definir as classes de problemas que a pesquisa irá trabalhar. Classes de problemas podem ser definidas como um conjunto de problemas práticos e teóricos que contenham artefatos úteis para aplicação nas organizações (Dresch, 2013). As classes de problemas da pesquisa precisam estar relacionadas ao seu objetivo geral que é: propor um modelo para governança de dados para o Observatório. E aos seus objetivos específicos que são:

identificar técnicas, ferramentas e mapear boas práticas de governança de dados; mapear o processo atual de governança de dados; identificar os elementos de governança de dados presentes em modelos e *frameworks* para aplicação no Observatório e estabelecer as diretrizes para implementação do modelo conceitual para o Observatório.

Com isso, por meio da revisão da literatura, foi possível compreender melhor as classes de problemas e apresentar artefatos existentes que apoiaram no desenvolvimento de recomendações para a governança de dados do Observatório. As classes de problemas identificados são relacionadas, principalmente, à qualidade de dados, segurança de dados e definições de papéis e responsabilidade que, de alguma forma, estão relacionados e acabam prejudicando a utilização dos dados por parte da organização. Além disso, foi possível encontrar modelos e *frameworks* provenientes de literatura acadêmica com os mais diversos objetivos, desde *frameworks* aplicados a contextos específicos de uma organização, como a utilização de dados em um hospital, passando por modelos e *frameworks* mais gerais, trazendo uma revisão sistemática da literatura, até artefatos mais voltados a setores, como a proposição de um *framework* de governança de dados para indústria 4.0 e inteligência artificial (Janssen *et al*, 2020; Jones, 2018; Thomas, 2009; Zorrilla; Yebenes, 2022).

O objetivo dessa etapa é relacionar os artefatos encontrados com as classes de problemas definidas pela pesquisa.

3.1.5 Proposição de artefatos

Com as classes de problemas e os artefatos existentes identificados e justificados de acordo com a metodologia proposta, o pesquisador iniciou a quinta etapa da *DSR*, que foi a proposição dos artefatos para resolver um problema específico (Dresch, 2013).

Essa etapa é necessária, pois a identificação das classes de problemas e artefatos existentes é apenas o mapeamento dos possíveis artefatos que podem ser aplicados para resolver o problema. Porém, mesmo essas soluções existentes na literatura estando consolidadas, elas precisam ser adequadas ao ambiente estudado. Para isso, é importante ponderar o contexto interno da organização estudada, bem como a viabilidade de sua aplicação.

De acordo com Manson (2006), essa etapa é caracterizada como uma etapa de testes e processos de criação em que as soluções encontradas na etapa anterior interagem até que se atinja uma viabilidade razoável para aplicação.

No contexto da pesquisa, foi proposto um modelo com base na literatura e no conhecimento prático do autor. O modelo foi avaliado pelo gestor imediato, com o objetivo de verificar a adaptabilidade do modelo ao Observatório.

Inicialmente é preciso fundamentar a escolha pela proposição de um modelo e não um *framework*. De acordo com Macedo e Souza (2022) um *framework* conceitual deve ser configurado como uma estrutura lógica, que apresenta um conjunto de procedimentos que o pesquisador irá desempenhar, representando de forma visual como os conceitos utilizados no estudo se relacionam entre si. O modelo conceitual por sua vez, é caracterizado por demonstrar teorias de forma mais abstrata. Um modelo conceitual explica amplamente o fenômeno de interesse, expressa suposições e reflete uma postura filosófica (Bem; Coelho, 2014; Macedo; Souza, 2022). Na elaboração de um modelo conceitual são necessárias as seguintes etapas: observação, interpretação, transformação e verificação (Alves, 2006).

3.1.6 Avaliação do artefato

Com o artefato projetado e desenvolvido, faz-se necessária a sua avaliação, que deve ser rigorosamente executada (Hevner *et al.*, 2010). O objetivo desta etapa é observar e medir como o artefato está se comportando em relação à solução proposta para o problema identificado nas etapas anteriores (Dresch, 2013).

Apesar de existir uma etapa específica para a avaliação do artefato, é recomendado que essa avaliação seja feita durante a execução das outras etapas, com o objetivo de evitar retrabalhos. Hevner *et al.* (2010), argumenta que existem 5 formas para avaliação de um artefato: i) observacional; ii) analítico; iii) experimental; iv) teste; e, v) descritivo.

A avaliação **observacional** pode usar elementos de estudo de caso e de campo, como planejamento, coleta e análise de dados, e o relato final do pesquisador. Seu objetivo é analisar o comportamento do artefato em profundidade e em ambiente real (Hevner *et al.*, 2010). Métodos e técnicas **analíticas** também avaliam o artefato, focando em sua estrutura interna e interação com o ambiente externo, visando verificar seu desempenho e impacto no sistema (Hevner *et al.*, 2010).

A avaliação **experimental** é realizada por experimentos controlados ou simulações (Hevner *et al.*, 2010). A avaliação por **teste** inclui teste funcional (Black Box) e estrutural (White Box), focando no ponto de vista do usuário e na avaliação do sistema interno, respectivamente (Hevner *et al.*, 2010). A avaliação descritiva, por sua vez, analisa a utilidade do artefato, utilizando a literatura ou construindo cenários para demonstrá-la (Hevner *et al.*, 2010).

Além das formas propostas por Hevner *et al.* (2010), há outras maneiras de avaliar um artefato, como grupos focais, que apoiam tanto o desenvolvimento quanto a avaliação do artefato (Bruseberg; Mcdonagh-Philp, 2002). Em *Design Science Research*, grupos focais promovem discussões profundas e colaborativas sobre os artefatos (Dresch, 2013). Bruseberg e Mcdonagh-Philp (2002) sugerem combinar grupos focais com outras técnicas para apoiar discussões, facilitar a triangulação dos dados e gerar novas ideias.

Para este trabalho, foi feita a avaliação do artefato com o gestor do Observatório, com o objetivo de verificar a aderência e adaptabilidade do modelo ao Observatório. Foram apresentados os elementos do modelo juntamente com seus relacionamentos, justificando o porquê daqueles elementos estarem presentes no modelo, assim como identificar se o modelo está alinhado ao contexto em que ele será aplicado (Schneider; Schimitt, 1998)

3.1.7 Projeto e desenvolvimento do artefato

A partir da seleção dos artefatos realizada na etapa anterior, é feita a seleção de apenas um artefato para seguir com o desenvolvimento da pesquisa. O projeto deste artefato deve considerar todos os seus componentes de funcionamento, levando em conta suas limitações e suas relações com o contexto da organização em que a pesquisa será aplicada (Dresch, 2013). Esta etapa nada mais é do que um produto das etapas anteriores, por isso, é fundamental que o desenvolvimento do artefato esteja de acordo com as soluções propostas na etapa anterior e com os problemas e classes de problemas elencados durante as etapas da pesquisa.

No projeto do artefato selecionado, são descritos todos os procedimentos necessários para sua construção e sua avaliação.

Para o desenvolvimento do artefato na pesquisa, foram considerados diversos *frameworks* encontrados na literatura. A partir deles foram retirados elementos que se

encaixavam às necessidades do Observatório e, com isso, foram definidas as dimensões, o comitê de governança de dados e os princípios para o modelo. Para os fatores influenciadores, foi considerada a experiência do autor associada com elementos encontrados na literatura que impactam no bom funcionamento de um modelo ou *framework* de governança de dados.

3.1.8 Explicitação das aprendizagens

Para Dresch (2005) o objetivo desta etapa é assegurar que a pesquisa realizada sirva como produção de conhecimento, tanto prático, como teórico. Com isso, nesta etapa são descritos os sucessos e insucessos decorrentes de toda a pesquisa, de maneira que o leitor poderá utilizar o documento como um *benchmark* para futuras pesquisas.

O tema governança de dados é pouco conhecido e explorado, mesmo para profissionais mais familiarizados com dados. Isso dificultou a discussão de alguns tópicos do modelo com outros analistas. Também foi percebido que há muita confusão de conceitos relacionados a governança e, por isso, a etapa de conscientização do problema deveria ser mais didática.

Outra dificuldade foi encontrar artigos sobre o tema na literatura, já que existem muitos modelos e *frameworks* conceituais, mas com pouca aplicação prática, o que dificultou a utilização de alguns elementos para o modelo.

Como pontos positivos, destaca-se o aumento de todo o Observatório para conscientização da governança de dados que passou a ser um dos assuntos centrais da área.

3.1.9 Comunicação dos resultados

Nesta etapa são expostos os conteúdos resultantes desta pesquisa, que no caso é a dissertação de mestrado. Dresch (2013), ressalta que nesta etapa e na anterior podem surgir novos *insights* para o pesquisador e que isto poderá servir de subsídio para aprimorar a pesquisa nas etapas anteriores.

Para esse trabalho, a comunicação será feita a partir da publicação da dissertação em *web sites* que o público poderá acessar e realizar a leitura do trabalho.

4 PROPOSTA DE MODELO CONCEITUAL DE GOVERNANÇA DE DADOS PARA O OBSERVATÓRIO

A presente seção apresenta a compilação de ideias obtidas a partir do uso do raciocínio dedutivo realizado durante a análise da revisão de literatura baseados nos constructos: Governança de dados; Observatórios; Dados Abertos Governamentais; e, Frameworks e Modelos de Governança de dados. O raciocínio dedutivo transforma o conhecimento geral em um conhecimento específico, permitindo um aprofundamento nos argumentos. Como resultado emerge o modelo conceitual da governança de dados, em caráter preliminar, sendo que este é resultado dos estudos teóricos. A Seção 4.1 apresenta as justificativas para desenvolver um modelo conceitual. Na Seção 4.2 são descritos os pressupostos para o desenvolvimento do modelo. Na Seção 4.3 é apresentada a construção do modelo conceitual com seus respectivos elementos.

4.1 O PORQUÊ DO MODELO CONCEITUAL DE DADOS

A escolha pelo modelo conceitual e não por um *framework* se deve ao fato de um *framework* conceitual ser configurado como uma estrutura lógica, que apresenta um conjunto de procedimentos que o pesquisador irá desempenhar, de forma visual, como os conceitos utilizados no estudo se relacionam entre si (Macedo; Souza, 2022).

Um modelo conceitual é uma representação, seja ela escrita ou gráfica, que descreve os conceitos mais significativos que compõem um problema. Para Grove, Burns e Gray (2013), os modelos conceituais são exemplos de grandes teorias e são altamente abstratos com constructos relacionados. Ele relata suposições, teorias e se assemelha a uma postura filosófica. Um modelo conceitual pode ser considerado como a imagem de um evento (Macedo; Souza, 2022) e é constituído por quatro etapas: observação, interpretação, transformação e verificação (Alves, 2006).

A escolha do tema está dentro de um campo de especialização do autor. Já a literatura revelou diversas áreas e temáticas relacionadas à governança de dados que não eram consideradas pelo autor em seu trabalho anteriormente. O isolamento dos elementos importantes ocorreu por meio da descoberta de elementos específicos da literatura e a descoberta de como eles podem se relacionar. Assim, surge o modelo conceitual de governança de dados do Observatório.

4.2 PRESSUPOSTOS PARA A CONSTRUÇÃO DO MODELO CONCEITUAL

A partir dos elementos abordados no referencial teórico e das necessidades do Observatório em relação a governança de dados, foram estabelecidos cinco pressupostos que são os motivadores para a construção do modelo, com o objetivo de aumentar a maturidade do Observatório em governança de dados para que seus ativos em dados atendam os objetivos de negócio.

Pressuposto 1: melhorar a gestão dos metadados do *data lake* do Observatório, que possui uma grande quantidade de bases de dados extraídas de fontes públicas, sendo cada fonte com características distintas, periodicidade de atualização diferentes, além de fornecerem informações sobre diversas temáticas, como saúde, educação, economia etc., possibilitando a construção de diversos indicadores.

Pressuposto 2: minimizar as deficiências relacionadas à qualidade dos dados, tais como desatualização e inconsistências, por meio de um processo que assegure a entrega de dados de alta qualidade ao usuário. Dado que o Observatório se baseia principalmente em dados públicos, a maioria dos dados disponibilizados pelos produtos do Observatório pode ser verificada em outras plataformas e até mesmo na fonte oficial. Portanto, é crucial garantir a compatibilidade entre as informações fornecidas pelo Observatório e aquelas disponíveis, especialmente na fonte oficial.

Pressuposto 3: definir princípios claros e precisos para a utilização dos dados, considerando que a maioria deles está armazenada em nuvem. Qualquer uso inadequado dos dados, seja no armazenamento, processamento ou em outras atividades, pode resultar em custos significativos para o Observatório.

Pressuposto 4: garantir a segurança dos dados que são transacionados no Observatório, por meio de normas e regras que assegurem que os dados sensíveis utilizados pelo Observatório estejam em conformidade com as leis brasileiras e normas internacionais.

Pressuposto 5: associar os conjuntos de dados disponíveis a responsáveis, para que qualquer mudança nestes conjuntos tenha uma supervisão por algum analista que irá analisar essa mudança e adaptar aos processos internos da equipe de dados, para que se mantenha a qualidade dos dados.

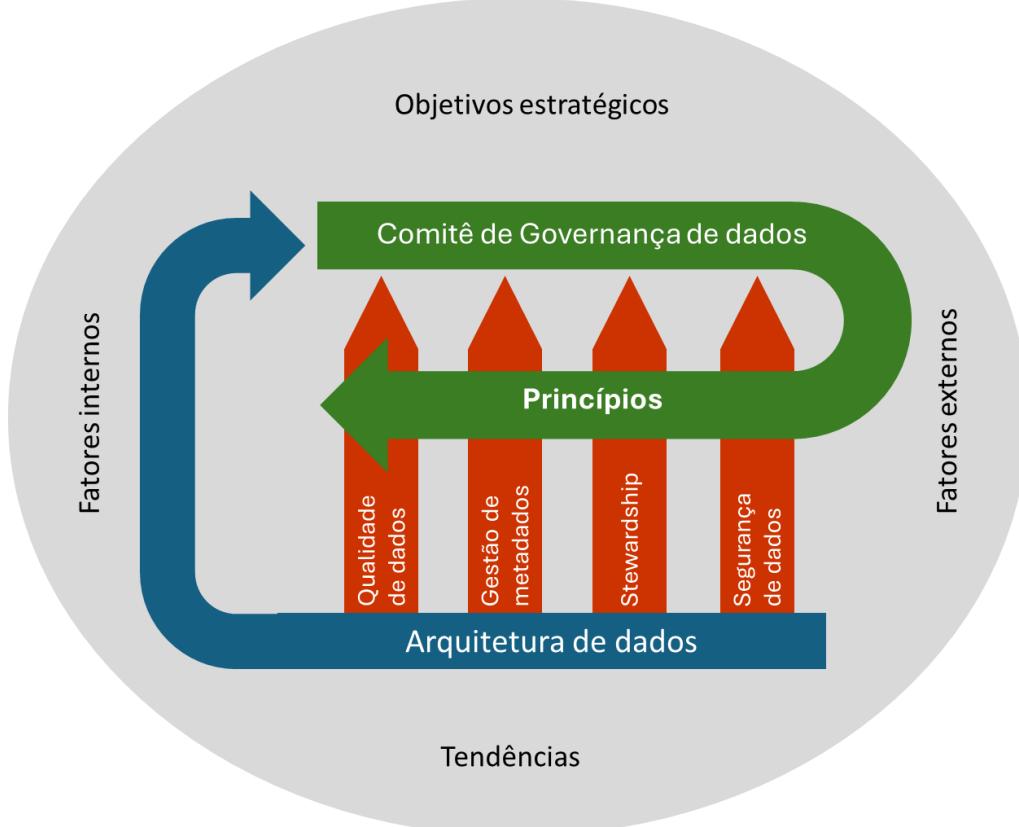
4.3 CONSTRUÇÃO DO MODELO CONCEITUAL

Para a construção do modelo conceitual, foram levados em consideração dimensões retiradas da literatura a respeito de governança de dados e a experiência prática do autor com o tema e com o contexto do Observatório. As principais referências utilizadas na literatura foram o DAMA (2017) e DGI (2021). Elas possuem dimensões que são agrupamentos de diversos elementos, que corroboram com as áreas de conhecimento da governança de dados e devem estar alinhadas com os objetivos estratégicos da organização, bem como com as boas práticas de governança de dados disponíveis na literatura.

O modelo conceitual foi construído com o objetivo de não apenas facilitar a interação entre as partes envolvidas, mas também estabelecer responsabilidades. Como forma de sintetizar a revisão bibliográfica, surgiu o modelo conceitual que engloba os conceitos de governança de dados e gestão de dados. Para a sua construção e seleção das dimensões, além da literatura, foi levado em conta o que o Observatório já possui de práticas de governança de dados e as suas necessidades, que foram apontadas nos pressupostos.

Com base nos pressupostos, foi elaborado o modelo conceitual que se propõe a desenvolver a governança de dados no Observatório, com o objetivo de nortear um programa de governança de dados. Neste sentido, o modelo conceitual é apresentado na Figura 2.

Figura 2 - Modelo conceitual de governança de dados do Observatório



Fonte: Elaborado pelo autor (2024)

Como é possível observar na Figura 2, o modelo possui cinco **dimensões** (qualidade de dados, gestão de metadados, *stewardship*, segurança de dados e arquitetura de dados) que são gerenciados a partir de **princípios** que são definidos pelo **comitê de governança de dados**. Todos esses componentes são influenciados por **fatores** que podem ter impacto no funcionamento do modelo, visto que, ele irá operar em um ecossistema complexo que exige mudanças e adaptações constantes. A base deste modelo é a **arquitetura de dados**, que fornecerá todas as ferramentas necessárias para otimizar a aplicação dos princípios para cada dimensão e auxiliará o comitê na gestão da governança de dados no Observatório, verificando se os princípios estão sendo respeitados.

Para a coordenação da governança de dados no Observatório e manutenção da sua sustentabilidade, é fundamental a criação de um comitê multidisciplinar que vai definir os princípios para cada dimensão e direcionar o programa de governança de dados para as estratégias de negócio estabelecidas para o Observatório. Os princípios têm a função de direcionar a governança de dados no Observatório, e são necessários princípios que nortearão outras dimensões sugeridas nesta pesquisa.

Portanto, essa dimensão é transversal a todas as outras. Além disso, os princípios poderão guiar o Observatório para mudança de conduta e cultura, buscando aumentar a literacia em dados da equipe do Observatório (Khatri; Brown, 2010; DAMA, 2017; Zorrilla; Yebenes, 2022).

As dimensões consideradas para o modelo foram definidas a partir de *frameworks* da literatura e das necessidades do Observatório. Cada uma das cinco dimensões tem seu papel específico que se relaciona a outras dimensões e que são direcionadas a partir dos princípios definidos pelo comitê. A qualidade de dados visa estabelecer processos para garantir a qualidade de dados, assim como métricas de acompanhamento do nível de qualidade dos dados gerenciados pelo Observatório (Mashoufi *et al.*, 2023; Wook *et al.*, 2021; DAMA, 2017; DGI 2021; Khatri; Brown, 2010). Já a gestão de metadados tem o objetivo de facilitar a organização e compreensão dos dados disponíveis, por meio da criação de glossários e catálogos que ajudarão a entender e dar contexto aos dados, além de disseminar os conhecimentos sobre os dados disponíveis (DAMA, 2017; DGI, 2021; Zorrilla; Yebenes, 2022; Barbieri, 2020). A dimensão *stewardship* é fundamental para operacionalizar o modelo, definindo a responsabilidade para as atividades de gerenciamento de dados, que são de responsabilidade de contribuidores individuais, bem como responsabilidades para grupos de administradores de dados (DAMA, 2017; DGI, 2021; Khatri; Brown, 2010). Outra dimensão fundamental é a segurança de dados que desenvolve controles e práticas com o objetivo de garantir a segurança e a privacidade necessária aos dados utilizados, além de assegurar que os dados sensíveis estão seguindo leis, normas e regulamentações impostas pelo governo (DAMA, 2017; Zorrilla; Yebenes, 2022). A dimensão base de todas as outras dimensões é a arquitetura de dados, que define ferramentas e seus inter-relacionamentos com o objetivo de atender a estratégia da governança de dados. É a dimensão base para todas as outras, pois otimiza os processos e auxilia o controle dos dados (DAMA, 2017).

Os fatores influenciadores são fundamentais para garantir o bom funcionamento e a sustentabilidade do modelo. Portanto, é essencial mapear os fatores que podem impactar seu desempenho do modelo. Eles podem surgir de aspectos legais que variam conforme as definições dos governos estaduais e federais, além de normas e regulamentações institucionais que afetam o orçamento da área e sua função dentro da instituição. Ademais, no setor de TIC, as mudanças são

frequentes, com novas ferramentas tecnológicas surgindo quase diariamente. Portanto, o modelo deve considerar esses fatores influenciadores, visto que o comitê precisa estar atento a essas novas ferramentas para aumentar a eficiência dos processos relacionados aos dados, desde o armazenamento e tratamento de dados até a catalogação de metadados. Por fim, o último e talvez mais importante fator influenciador são os objetivos estratégicos. Este modelo está sendo desenvolvido justamente para alinhar os ativos de dados com esses objetivos, sendo essa uma das principais funções de qualquer modelo ou *framework* (Zorrilla; Yebenes, 2022).

Para descrever de forma mais aprofundada cada um dos elementos definidos para o modelo conceitual e relacioná-los com os pressupostos estabelecidos e com os desafios do Observatório, os próximos tópicos tratam disso.

4.3.1 Comitê de governança de dados

O comitê de governança de dados é fundamental para garantir a manutenção do programa de governança de dados na organização. Ele irá estabelecer, monitorar e promover as melhores práticas de gestão de informações e manter atualizada a política de governança de dados. O comitê precisa ser multidisciplinar, com profissionais com poder de decisões nas áreas do Observatório, para adequar a gestão e a governança de dados aos seus interesses. Além disso, precisa fornecer suporte técnico aos responsáveis pelos dados, facilitar a capacitação e assegurar a transferência de conhecimento, enquanto dissemina as melhores práticas na gestão de dados entre os *stakeholders*. Também será responsável pela gestão do catálogo de dados e da plataforma de qualidade de dados, além de oferecer suporte aos processos de captação e integração de dados (Silva, 2023).

4.3.2 Princípios

Os princípios estabelecidos para a governança de dados possuem uma relação com cada dimensão definida para o modelo conceitual e devem estar de acordo com as estratégias de negócio do Observatório (Zorrilla; Yebenes, 2022). Eles são responsáveis por definir os comportamentos necessários para o uso dos dados. Por exemplo, em uma gestão de metadados, os princípios auxiliam a estabelecer o que, como e quando um item de dados deve ser inserido em um catálogo de metadados,

bem como definir os processos para manutenção desse catálogo. Além disso, os dados são utilizados tanto para fins institucionais como para fins comerciais. Para tanto, é necessário que haja princípios relacionados à qualidade de dados, como definição de padrões de tratamento, armazenamento e processamento, que garantam a qualidade do dado que é repassado ao demandante. Portanto, cada princípio é apoiado por uma justificativa e um conjunto de implicações, estando alinhados aos objetivos de negócio da instituição (Khatri; Brown, 2010).

4.3.3 Qualidade de dados

Devido ao número de bases de dados abertos governamentais trabalhadas no Observatório e considerando que cada uma possui uma especificidade diferente, é fundamental que padrões de qualidade sejam definidos e replicados a todos os ativos de dados. Além disso, é importante ter métricas de avaliação da qualidade para que se possa medir a qualidade dos dados ao longo do tempo e tomar ações, caso necessário.

A qualidade de dados possui constructos multidimensionais que refletem a acuracidade, confiança, completude, consistência, temporalidade e relevância do dado (Mashoufi *et al.*, 2023; Wook *et al.*, 2021; Khatri; Brown, 2010). Os principais desafios para manter a qualidade dos dados são: velocidade e volume dos dados, dados armazenados em silos, complexidade de integração dos dados, erro humano, governança de dados e segurança (Pansara, 2023). No contexto do Observatório, que trabalha majoritariamente com dados abertos governamentais, esses desafios são muito presentes no cotidiano dos analistas, já que não existe um padrão único para os dados, ou seja, cada base tem suas especificidades em relação a forma de extração, estrutura de tabela e periodicidade de atualização.

A linhagem de dados é outro elemento importante na qualidade de dados. Por meio dela é possível monitorar por onde os dados passam até chegarem ao seu destino, que geralmente é uma visualização. Caso algum erro no dado ocorra, fica facilitada a localização do processo que contribuiu para que o dado ficasse incorreto. A rastreabilidade dos dados, proporcionada pela linhagem, facilita o mapeamento dos termos de negócios disponíveis no dicionário de dados para objetos como tabelas, registros e campos.

Essa dimensão atende o segundo pressuposto para o desenvolvimento do modelo conceitual, auxiliando o Observatório a elevar a qualidade dos seus dados.

4.3.4 Gestão de metadados

Associado ao primeiro pressuposto, é fundamental estabelecer uma gestão de metadados que irá auxiliar o Observatório a registrar o conhecimento a respeito dos dados que tem disponíveis e que faz uso.

Os metadados (dado sobre dado) auxiliam no aumento da eficiência dos processos de atendimento de demandas, visto que reduz significativamente a procura pelo dado necessário (DAMA, 2017).

Através de uma organização, diferentes indivíduos terão diferentes níveis de conhecimento em dados, porém, nenhum indivíduo vai saber sobre todas as bases de dados trabalhados. Esta informação necessita ser documentada para evitar riscos de perda de conhecimento. Os metadados fornecem o significado inicial e a gestão para organizar o conhecimento da organização sobre os dados (DAMA, 2017). Entretanto, a gestão de metadados não é apenas um desafio de gestão de conhecimento, mas também um grande desafio para a gestão de riscos.

Os metadados tem a função de dar contexto aos dados, um é dependente do outro, são elementos de suporte na transformação do dado em informação e de informação para conhecimento. Alinhado com a comunicação, ajuda a equipe a aumentar a literacia sobre os dados disponíveis na organização (Barbieri, 2020).

4.3.5 *Stewardship*

Como já citado anteriormente, a grande quantidade de bases de dados no Observatório dificulta a gestão da qualidade dos dados e o monitoramento do seu ciclo de vida. Com isso, a definição de responsabilidade dentro da equipe para cada base é importante para garantir que o dado que está chegando no usuário final seja confiável e que ele não se perca entre as tecnologias pelas quais o dado percorre. Isso é um dos motivadores para a criação desse modelo conceitual. Portanto, essa dimensão se relaciona com o quinto pressuposto já apresentado.

O foco inicial do *stewardship* é definir as especificidades do negócio e os seus valores bases para cada área e subárea (DAMA, 2017). Além disso, em alguns casos,

os responsáveis pelos dados (*stewards*) definem e mantêm os requisitos de qualidade de dados e regras de negócios para aplicar aos dados. Também, ajudam a identificar e resolver questões relacionadas a dados e auxiliam a desenvolver e manter as políticas e procedimentos de dados (Khatri; Brown, 2010).

4.3.6 Segurança de dados

A dimensão segurança de dados é fundamental para o modelo conceitual, visto que mudanças nas leis e nos procedimentos a serem adotados com dados sensíveis ocorrem frequentemente. O impacto por não seguir uma lei ou procedimento pode ser crucial para a estratégia da organização.

Além disso, este elemento deve considerar os *stakeholders* envolvidos nos processos de dados, seja um *stakeholder* interno ou externo, como clientes e fornecedores.

Assim, nesta dimensão são tratados procedimentos para assegurar que as políticas de governança de dados estejam em conformidade com os aspectos regulatórios, com diretrizes claras de: por quem, para que e como os dados sensíveis disponíveis serão utilizados.

Essa dimensão visa proteger os ativos de dados em alinhamento com as regulações relativas à privacidade e confidencialidade, acordos contratuais e os requisitos de negócios (DAMA, 2017). Para isso, é de responsabilidade da governança de dados controlar acessos a dados corporativos para garantir que apenas as pessoas que devem ter acesso possam utilizá-los; estabelecer requisitos de acesso a dados para *stakeholders*, sejam eles internos ou externos, como clientes e fornecedores; além de monitorar o cumprimento das políticas estabelecidas e estar sempre atualizado quanto às regulamentações envolvendo os dados, tanto no âmbito nacional, como internacional.

4.3.7 Arquitetura de dados

Essa dimensão, de alguma forma, atende todas as outras e é fundamental para o funcionamento do modelo já que os sistemas e as tecnologias que armazenam ou processam os dados precisam estar em harmonia para suportar uma governança de

dados que preze pela segurança, qualidade e eficiência nos processos de dados. Tudo isso alinhado com a estratégia da organização.

Os elementos da arquitetura de dados incluem especificações utilizadas para descrever o estado atual, definir requerimentos de dados e controlar os ativos de dados de acordo com a sua estratégia (DAMA, 2017). A arquitetura de dados de uma organização é descrita pela integração de diferentes documentos de desenvolvimento, com diferentes níveis de abstrações, incluindo padrões de como os dados são coletados, armazenados, tratados, utilizados e descartados. Também é descrita por todas as camadas e caminhos que o dado percorre pelos sistemas da organização (DAMA, 2017).

4.3.8 Fatores influenciadores

Os fatores influenciadores desempenham um papel importante no funcionamento eficaz do modelo de governança de dados. Portanto, é essencial identificá-los e incorporá-los ao modelo, garantindo que sejam devidamente considerados, especialmente pelo comitê de governança de dados. O modelo desenvolvido contém quatro principais fatores influenciadores: fatores externos, fatores internos, tendências e objetivos estratégicos.

Os objetivos estratégicos se referem “onde” o Observatório quer chegar como área. Isso tem influência direta no comitê, cujos participantes orientam suas atividades de acordo com esses objetivos e por conseguinte terá influência no modelo e em todas suas dimensões, como exemplo, temos a rede de observatório que pode ter influência direta nos objetivos estratégicos e que pode impactar na dimensão arquitetura de dados, que pode ser compartilhada com outros Observatórios.

Os fatores externos se referem a quaisquer mudanças na legislação local, estadual e federal que podem afetar a forma como é feita a gestão de dados, principalmente a dimensão segurança de dados, já que restrições e regulamentações são atualizadas com recorrência. Além disso, como o Observatório possui produtos de dados, para cada produto existem concorrentes, o que é um elemento que pode ter algum tipo de influência, principalmente, na abrangência dos dados em relação às temáticas e na forma e na velocidade que eles ficam disponíveis para o usuário dos dados do Observatório. Por isso, o comitê precisa estar atento aos fatores externos,

para que a sustentabilidade do Observatório não seja afetada por falta de atenção a este tipo de fator.

Os fatores internos relacionam-se aos aspectos institucionais que influenciam o Observatório em sua estratégia e que acabam afetando o modo como é feita a gestão dos dados do Observatório, em todas as suas dimensões. Eles refletem as mudanças que acontecem na instituição, em relação ao seu modelo de gestão e aos objetivos que são estabelecidos para o Observatório e podem acabar mudando o seu funcionamento e o mercado em que atua.

A tendências se referem, principalmente, às tendências tecnológicas que mudam constantemente e com uma velocidade muito grande. Novas ferramentas para armazenamento, processamento e visualização de dados são criadas quase que diariamente. Por isso, o comitê de governança de dados precisa estar atento para ajustar não só a arquitetura de dados com novas ferramentas, mas as dimensões que podem tirar proveito dessas tendências tecnológicas.

No próximo tópico, serão elencados os mecanismos necessários para operacionalizar cada elemento do modelo.

4.4 MECANISMOS PARA OPERACIONALIZAR OS ELEMENTOS DO MODELO

Com as dimensões definidas, faz-se necessário incluir os mecanismos que estarão presentes em cada dimensão e que irão auxiliar a operacionalizar o modelo conceitual para o Observatório

De acordo com Moreira, Freire e Aguiar (2017), os mecanismos de governança são responsáveis por garantir que os processos de conhecimento, como compartilhamento, retenção e criação de conhecimento, sejam conduzidos por meio de procedimentos estabelecidos e modelos organizacionais. Esses mecanismos têm como objetivo principal sustentar a estratégia da organização. As empresas geralmente projetam uma série de mecanismos para garantir a boa implementação de ações estratégicas relacionadas à GD. Esses mecanismos incluem estruturas que conectam dados ao negócio, procedimentos e normas para implantação de GD (Zhang; Sun; Zhang, 2022).

Os mecanismos de governança compreendem estruturas formais que conectam funções de negócios, TI e gerenciamento de dados, processos e procedimentos formais para a tomada de decisão e o monitoramento e práticas que

apoiam a participação ativa e a colaboração entre as partes interessadas (Abraham; Schneider; Vom Brocke, 2019).

No Quadro 3, são explicitados os mecanismos definidos para cada elemento do modelo.

Quadro 2 - Mecanismos dos elementos do modelo

Elementos de governança de dados	Mecanismos de governança de dados	Fonte
Comitê de governança de dados	<p>Desenvolver e manter a estratégia para o programa de governança de dados</p> <p>Estabelecer canais de comunicação eficientes para a governança de dados</p> <p>Monitorar os custos relacionados a dados</p> <p>Definir princípios para cada dimensão do programa de governança de dados</p> <p>Realizar interações com stakeholders institucionais</p>	Alhassan, Sammon, Daly (2018); Barbieri (2020); DAMA (2017); DGI (2021); Khatri, Brown (2010).
Qualidade de dados	<p>Definir padrões para qualidade de dados</p> <p>Definir requisitos para os dados</p> <p>Definir indicadores de monitoramento da qualidade dos dados</p> <p>Definir políticas de armazenagem de dados</p> <p>Gerenciar o versionamento e a integração dos modelos de dados</p> <p>Implementar monitoramento do ciclo de vida dos dados</p> <p>Implementar processos para inventariar os dados</p> <p>Definir processos para melhoria da qualidade dos dados</p> <p>Implementar linhagem de dados</p>	Abraham, Schneider, Vom Brocke, (2019); Alhassan, Sammon, Daly (2018); Al-Ruithe, Benkhelifa, (2017); Attard (2015); Barbieri (2020); Bento, Neto, Côrte-Real, (2022); DAMA (2017); DGI (2021); Kim, Cho (2017); Ladley, (2019); Zorrilla, Yebenes (2022).
Gestão de metadados	<p>Definir arquitetura de metadados (centralizada ou descentralizada)</p> <p>Definir um glossário para os dados</p> <p>Implementar um catálogo de metadados</p> <p>Implementar dicionário de dados</p> <p>Definir responsáveis por manter e atualizar o catálogo</p>	Alhassan, Sammon, Daly (2018); Barbieri (2020); DAMA (2017); DGI (2021); Khatri, Brown (2010); Jones (2018).
Stewardship	Definir responsáveis por cada conjunto de dados	

	Estabelecer os papéis e responsabilidades de cada <i>data steward</i>	DAMA (2017); DGI (2021); Khatri; Brown, (2010); Jones (2018); Zhang; Sun; Zhang (2022).
Segurança de dados	Implementar controle de acesso a dados	
	Definir e monitorar padrões para os conjuntos de dados	
	Realizar a gestão da documentação	
	Implementar processos para a segurança de dados	DAMA (2017); Khatri, Brown (2010); Zhang; Sun; Zhang (2022).
	Implementar auditoria	
	Definir diretrizes para compartilhamento de dados	
	Definir classificação de grau de confidencialidade das informações	
	Implementar processo para monitorar a conformidade dos dados	
	Implementar processos para proteção de dados pessoais	
Arquitetura de dados	Definir arquitetura de dados	Abraham, Schneider, Vom Brocke, (2019); Alhassan, Sammon, Daly (2018); Al-Ruithe, Benkhelifa (2017); DAMA (2017); Silva 2023); Soares (2012); Zorrilla, Yebenes (2022)
	Implementar processo de <i>data warehousing</i>	
	Definir infraestrutura de dados	
	Definir investimentos em TIC	
	Monitorar performance das atividades de processamento de dados	

Fonte: Elaborado pelo autor (2024)

4.5 IMPLEMENTAÇÃO DO MODELO

Para a implementação do modelo conceitual proposto neste trabalho, inicialmente é necessário avaliar o que já foi construído e o que é praticado pela área de dados no Observatório. Assim, no Quadro 3, apresenta-se essa avaliação, a partir das dimensões definidas para o modelo, com o cenário atual e a recomendação para a área, considerando o modelo conceitual proposto com os mecanismos apresentados no Quadro 2.

Quadro 3 - Implementação do modelo conceitual

Comitê de governança de dados	
<p>O primeiro passo para a estruturação da governança de dados deve ser a criação do comitê, ele deve ser composto por todos os colaboradores que sejam exerçam um papel de líder de seus times e que de alguma forma utilizam os dados no Observatório, além disso ele deve considerar a participação eventual de outros <i>stakeholders</i> como gestores da área de segurança de informação da instituição, gestores de TI, entre outros. Com a equipe formada, o comitê pode se utilizar dos mecanismos acima definidos, para estabelecer a estratégia para governança de dados, com a visão, a missão, o que se espera de resultado com a governança e quem serão os seus beneficiários. É importante estabelecer uma comunicação com toda a equipe da área para disseminar o trabalho do comitê, divulgar os princípios estabelecidos e conscientizar a equipe sobre a importância da governança de dados. Outro importante ponto que foi mencionado no Quadro 2 é o monitoramento dos custos relacionados a infraestrutura necessária para os dados do Observatório, o comitê precisa constantemente verificar se o valor investido está sendo convertido em benefício para a área, prezando pela eficiência nos gastos.</p>	
Dimensão Qualidade de dados	
Cenário atual	<p>Existe um processo de atualização e validação dos ativos de dados presentes nas visualizações do Observatório, com definição de responsáveis para cada conjunto de dados. Além disso, existem três camadas de tratamento dos dados no <i>Big Data</i>, a <i>landing</i>, onde o dado bruto é extraído da sua fonte original e chega no <i>Data Lake</i>, a <i>silver</i>, onde o dado bruto é tratado, com a remoção de valores nulos, caracteres especiais e dados incorretos, e a <i>gold</i>, onde são aplicadas as regras de negócio do Observatório nos dados tratados da <i>silver</i>. O analista responsável pela atualização precisa acompanhar todo esse processo até o dado da <i>gold</i> chegar na visualização, sendo responsável também por monitorar quando os dados são atualizados na fonte, para detectar possíveis falhas de atualização no <i>Big Data</i>.</p>
Recomendação	<p>Tendo em vista os mecanismos apresentados no Quadro 2, é necessário definir padrões de estruturas de tabela, nomenclatura de colunas e tabelas, definições de esquemas de dados e formatos de tabelas que atendam os objetivos estratégicos da área. Para isso, a arquitetura será fundamental para definir ferramentas que atendam essas necessidades e consigam auxiliar na criação de indicadores de qualidade de dados. Além disso, o comitê deve definir locais de armazenamento de tabelas de acordo com o objetivo de cada tabela, considerando questões de custos e eficiência de processamento. Outro ponto importante é a gestão de ciclo de vida dos ativos de dados, com definições de critérios de perenidade das tabelas e do porquê elas precisam ser armazenadas e por quanto tempo, ou do contrário, porque a tabela deve ser eliminada. Também deve ser definido um processo de monitoramento e melhoria da qualidade de dados, a partir de treinamentos de colaboradores responsáveis pelos conjuntos de dados e regras claras de tratamento de dados. O monitoramento ficará facilitado com a linhagem de dados implementada em todos os objetos de dados da área. Assim, qualquer intercorrência que possa levar a um dado incorreto no destino, pode ser mais facilmente mapeado com a implementação da linhagem de dados, conforme sugerido no Quadro 2. Além disso, precisa haver uma</p>

	reavaliação no processo atual do <i>Big Data</i> , em que os dados passam pelas camadas de tratamento (<i>landing, silver e gold</i>), com o objetivo de adequar os princípios estabelecidos pelo comitê e verificar se este processo está de acordo com os objetivos estratégicos da área. Após estas etapas, é necessário definir os requisitos de qualidade de dados, levando em consideração os processos críticos por onde os dados passam, as regras de negócio e o impacto da qualidade do dado nos produtos do Observatório. Todos os processos necessitam de um versionamento e acompanhamento do que foi feito e o que está sendo feito com os dados, com a possibilidade de retornar a procedimentos adotados anteriormente, caso seja identificada alguma inconsistência no processo atual.
--	---

Dimensão Gestão de metadados

Cenário atual	Processo manual de construção e manutenção do catálogo de metadados e ausência de formalização de responsabilidades quanto a cada conjunto de metadados.
Recomendação	A arquitetura de dados é essencial para diminuir o esforço manual na gestão dos metadados, além de auxiliar na definição das responsabilidades de monitoramento dos catálogos de metadados com o objetivo de manter ele sempre atualizado. Também é fundamental definir o tipo de gestão de metadados, se vai ser descentralizado ou centralizado, considerando o tamanho da equipe do Observatório, recomenda-se centralizado, com um padrão único de dicionário de dados e catalogação dos metadados, possibilitando a integração destes objetos de metadados, devido a um padrão unificado. Como mencionado no Quadro 2, é fundamental definir responsabilidades para a manutenção dos metadados, caso contrário, eles ficaram desatualizados e não poderão ser utilizados pelos analistas em suas mais diversas funções.

Dimensão *Stewardship*

Cenário atual	Não existe um processo sistêmico de <i>stewardship</i> , ou seja, a definição de responsabilidades, porém, existe uma iniciativa no processo de qualidade de dados, em que são definidas responsabilidades como já explicado anteriormente.
Recomendação	A dimensão de <i>stewardship</i> auxilia outras dimensões do modelo na definição de papéis e responsabilidades por conjuntos de dados, o que também auxiliará no controle de acesso a dados, de acordo com os princípios definidos na segurança de dados, e quem ficará responsável pela catalogação dos metadados. Para isso, recomenda-se a definição dos <i>data stewards</i> , que serão separados em conjuntos de dados definidos pelo comitê de governança de dados do Observatório. Eles serão responsáveis, além da gestão dos acessos, também por todos os processos de qualidade e gestão de metadados dos seus respectivos conjuntos de dados. Outro fator importante, apontado no Quadro 2 como um mecanismo, é a documentação dos processos pelos <i>Data Stewards</i> , a fim de explicitar e compartilhar o conhecimento para realização dos processos relacionados a dados.

Dimensão Segurança de dados	
Cenário atual	O Observatório trabalha majoritariamente com dados públicos. Não existe utilização de dados sensíveis às organizações e dados pessoais que são resguardados pela LGPD.
Recomendação	Apesar de a área não possuir dados sensíveis, faz-se necessário implementar processos que visam garantir a segurança do dado que está sendo transacionado no ambiente do Observatório, seja através de controle de acessos, criação de ambientes privados, até auditorias com o objetivo de garantir que não haja falhas em termos de segurança e que as práticas de tratamento de dados do Observatório estejam de acordo com as leis vigentes. Estes processos devem ser definidos juntamente com a área de segurança da informação e TI da instituição, com o objetivo de evitar qualquer descumprimento das regras estabelecidas. O comitê tem um papel fundamental na articulação com outras áreas da, além da construção dos princípios, observando a legislação vigente e as normas da instituição, relacionadas ao tema. Também, é necessário gerenciar as permissões de acessos a dados junto aos <i>data stewards</i> , para garantir a privacidade e a confidencialidade dos dados, quando necessária, e assegurar a uniformidade de regras entre eles. Outra prática essencial para a segurança de dados do Observatório é a classificação do grau de confidencialidade das informações, definindo hierarquias como por exemplo, informações de acesso geral, informações somente para uso interno, informações confidenciais, informações confidenciais restritas etc.
Dimensão Arquitetura de dados	
Cenário atual	A área já tem um conjunto de ferramentas para trabalhar com dados de uma das principais nuvens do mercado. Essas ferramentas foram orquestradas junto a um fornecedor, com o objetivo de atender as demandas do Observatório com dados abertos governamentais.
Recomendação	Como a área precisa, muitas vezes, utilizar microdados para sua análise, é necessário um sistema de armazenamento de dados robusto, com ferramentas capazes de trabalhar com esses dados que vem de forma distinta e que precisam de atualização constante, principalmente mensais e semanais, o que acaba exigindo da arquitetura uma capacidade de construção de <i>pipelines</i> , ou seja, processamentos de dados conectados, sendo cada um dos processamentos tendo uma função, seja de extração, tratamento etc. Além disso, devem ser contempladas ferramentas de governança de dados disponíveis no mercado para auxiliar no gerenciamento pelo comitê do ambiente de dados do Observatório, conforme apontado no Quadro 2, sobre os mecanismos para as dimensões do modelo.
Fatores influenciadores	

Os fatores influenciadores devem ser observados principalmente pelo comitê, para que possa agir rapidamente na redefinição dos princípios das dimensões, pois estes fatores podem comprometer a sustentabilidade do programa de governança de dados. Além disso, o comitê terá que criar indicadores de monitoramento para o programa, além de se aliar a equipe de comunicação do Observatório, para que os princípios sejam frequentemente disseminados, compartilhados, lembrados e esclarecidos aos colaboradores usuários de dados e assim, aos poucos, incorporar a cultura de governança a todo time do Observatório.

Fonte: Elaborado pelo autor (2024)

Como já citado no tópico 4.3, a arquitetura é a base do modelo, sem ela a sua aplicação fica dificultada. Por isso a escolha das ferramentas é tão importante para a sustentabilidade do modelo e financeira do programa de governança do Observatório. Os fatores influenciadores já são observados atualmente, mas de forma desestruturada. Com o modelo, espera-se que o comitê estruture o processo de monitoramento dos possíveis fatores que podem afetar a governança de dados. As dimensões podem não ser aderidas de forma completa no Observatório, pois isso pode variar de acordo com decisões de gestão, mas os fatores influenciadores devem ser levados em conta para cada decisão que é tomada.

4.6 AVALIAÇÃO DO MODELO CONCEITUAL PROPOSTO

Para a verificação da aderência do modelo ao Observatório, faz-se necessária a avaliação do gestor da área. Assim, foi apresentado a ele todo o trabalho desenvolvido pelo autor desta pesquisa para chegar no modelo. Foi explicado o modelo, as suas dimensões, os fatores influenciadores e como esses elementos se relacionam.

Foi realizada uma entrevista presencial com o gerente de aproximadamente uma hora, com perguntas semiestruturadas e com perguntas gerais, para dar liberdade ao entrevistado nas suas respostas, sem perder o objetivo da entrevista que era a avaliação do modelo conceitual desenvolvido para o Observatório.

Primeiramente, foi apresentado o objetivo da entrevista e como ela seria utilizada no presente trabalho. Na sequência, foi feito a explicação do modelo com o seu detalhamento e sanadas algumas dúvidas dos gerentes. Por fim, foram realizadas as perguntas necessárias para a avaliação do modelo. As perguntas com suas respectivas respostas são apresentadas a seguir.

- a) Você considera a governança de dados importante para o Observatório? Por quê? O modelo é importante porque ele foca em elementos cruciais para um centro de inteligência, tais como: segurança, confiabilidade, qualidade, conformidade e gestão de riscos.
- b) As dimensões do modelo (qualidade de dados, segurança de dados, stewardship e gestão de metadados) são suficientes para um programa de governança de dados do Observatório? Sim.
- c) Os fatores influenciadores são suficientes para um programa de governança de dados do Observatório? Sim.
- d) Qual(is) outro(s) elemento(s) você traria para o modelo? Justifique sua resposta? Apesar das dimensões do modelo abrangerem com aderência os requisitos de governança de dados do Observatório, observa-se forte tendência na questão de integração de *data lakes*. Assim sendo, faz-se necessário contemplar, em um futuro próximo, tal dimensão.
- e) De maneira geral, o modelo conceitual é aderente ao Observatório? Sim, bastante aderente.
- f) Você considera que este modelo trará benefícios para o Observatório? Sim, pois vai garantir que o Observatório possa maximizar o valor de seus dados enquanto minimiza riscos e desafios associados.

Portanto, a partir da avaliação do gestor do Observatório, considera-se o modelo conceitual validado, já que não foi demonstrado nenhuma inconsistência e falta de aderência do modelo em relação ao contexto em que será aplicado.

5 CONSIDERAÇÕES FINAIS

Na sociedade 5.0, a quantidade de dados disponíveis das mais diversas fontes e formatos é enorme. É crescente também, a quantidade de ferramentas e soluções para trabalhar e extrair valor destes dados. Para as organizações, estas oportunidades são desafiadoras. Por um lado, existe a concorrência que se torna mais eficiente ao utilizar os dados de maneira apropriada e por outro há uma oferta cada vez maior de tecnologias e de dados disponíveis para utilização. Ao mesmo tempo, profissionais no mercado especializados no tema são mais caros e disputados internacionalmente. Assim, as barreiras administrativas na gestão dos ativos de dados devem ser mitigadas pelas organizações, para que ela se mantenha competitiva frente aos concorrentes e consiga utilizar todos esses recursos sem comprometer o seu funcionamento e a sua sustentabilidade financeira.

Esta pesquisa concentrou a revisão da literatura na governança de dados e seus *frameworks* e modelos. Também foi realizada a revisão dos temas dados abertos governamentais e observatórios. Ficou evidenciado na literatura pesquisada, que governança de dados é um assunto multidisciplinar envolvendo todos os níveis de hierarquia de uma organização e ao longo dos anos, tem ganhado cada vez mais notoriedade e vem sendo utilizada como referência em outras governanças que tem surgido, como por exemplo, governança de inteligência artificial. Com o tempo, as organizações estão tomando consciência de que é necessária uma governança bem estruturada antes de aderir às novas tecnologias presentes no mercado, caso contrário, a má utilização dessas tecnologias pode acarretar prejuízos altos.

Além da literatura, o autor utilizou a sua experiência ao trabalhar na área de governança de dados há quase três anos, e de uma busca documental para mapear todos os processos de governança de dados que foram e são utilizados no Observatório, identificando ferramentas utilizadas, papéis e responsabilidades destes processos e analisando como eles se relacionam com o funcionamento de outros processos do Observatório.

Em relação aos objetivos da pesquisa, considera-se que eles foram alcançados.

O modelo conceitual foi elaborado considerando a revisão de literatura, a experiência do autor, a avaliação do modelo pelo gestor e a proposta de implementação no Observatório. Dessa forma, considera-se que o objetivo geral, de

propor um modelo conceitual de governança de dados para o Observatório foi atingido.

No tocante ao objetivo específico **identificar técnicas, ferramentas e boas práticas para governança de dados**, foram identificados diversos modelos e frameworks presentes na literatura, desde os mais específicos feitos para indústria 4.0 e *cloud*, até os mais genéricos como o DAMA e o DGI. A partir deles foi possível encontrar ferramentas e boas práticas que foram utilizados no modelo proposto neste trabalho.

Em relação ao objetivo específico **mapear o processo atual de governança de dados do Observatório**, verificou-se que existem práticas que são adotadas e que são seguidas pelos colaboradores, porém falta uma visão sistêmica destas práticas e um alinhamento com os objetivos estratégicos da área e da instituição em que o Observatório está inserido.

Em relação ao objetivo **de identificar os elementos de governança de dados presentes em modelos e frameworks para aplicação no Observatório** buscou-se selecionar elementos da literatura com aderência a dinâmica de trabalho da área e também elementos que podem se encaixar aos processos já estabelecidos, assim como as características da área.

Sobre o objetivo **estabelecer as diretrizes para implementação do modelo conceitual para o Observatório** buscou-se olhar para o processo atual e, a partir dele, sugerir a implementação de acordo com cada dimensão, considerando os mecanismos de cada dimensão, mas sempre tendo em vista que todo esse processo precisa ser gerenciado pelo comitê de governança de dados.

Para avaliar o modelo proposto, foi apresentado o modelo ao gestor do Observatório para que ele, a partir da sua visão sistêmica da área e das suas estratégias para o Observatório, pudesse verificar a sua aderência e o que poderia contribuir não apenas para a área de dados, mas também para as outras áreas do Observatório, como produtos, projetos e economia, a fim de verificar o seu impacto nestas áreas e na sua contribuição para o negócio como um todo.

Com os objetivos específicos detalhados acima, foi possível responder à pergunta da pesquisa: **De que forma e a partir de quais elementos pode ser estruturado um modelo conceitual para a governança de dados do Observatório?** Por meio de metodologia adequada que buscou a intersecção da governança de dados ao contexto de Observatórios de indústria. Assim foi possível

alcançar o objetivo geral que foi de **propor um modelo conceitual de governança de dados para o Observatório**.

Cumpre destacar que esta pesquisa buscou relacionar a governança de dados aplicada ao Observatório, associando a literatura encontrada com aspectos empíricos do autor em seu trabalho diário. Com isso, foram definidas cinco dimensões para o modelo conceitual proposto de acordo com o que já era praticado no Observatório e levando em consideração as suas necessidades, de acordo com os seus objetivos estratégicos.

Como contribuições teóricas, o trabalho estabeleceu uma relação entre governança de dados, observatório de indústria e dados abertos governamentais abertos, apresentando modelos, *frameworks* e mecanismos de governança de dados que são mais aderentes aos outros temas da dissertação. Além disso, esta dissertação contribui para auxiliar a disseminação do tema de governança de dados e sua importância no momento atual, de popularização do uso de dados para tomada de decisões nas organizações.

Em relação às contribuições práticas, a dissertação apresentou como um modelo conceitual pode ser aplicado na prática, considerando as limitações de uma organização em seus recursos tecnológicos e humanos e a flexibilidade que um programa de governança de dados precisa ter para ser sustentável e relevante para a organização, podendo ser utilizado como um guia para possíveis aplicações em outras organizações. Nesse caso, levando em consideração suas características e sendo, portanto, necessário ajustes no modelo desenvolvido.

Como sugestão para estudos futuros, recomenda-se a análise de outros observatórios de indústrias em relação ao modelo desenvolvido e, se for o caso, até a proposição de um *framework* que poderá servir de referência para a construção de um Programa de governança de dados. Além disso, como contribuição adicional sugere-se a validação do modelo desenvolvido com especialistas de governança de dados de organizações de outra natureza e a validação com gestores de outros observatórios de indústrias.

REFERÊNCIAS

- ABRAHAM, Rene; SCHNEIDER, Johannes; VOM BROCKE, Jan. Data governance: a conceptual framework, structured review, and research agenda. **International journal of information management**, v. 49, p. 424-438, 2019.
- AHMAD, Wan Azlin Zurita Wan; MUKHTAR, Muriati; YAHYA, Yazrina. Validating the contents of a social content management *Framework*. In: **2017 6th International Conference on Electrical Engineering and Informatics (ICEEI)**. IEEE, 2017. p. 1-6.
- AL-BADI, Ali; TARHINI, Ali; KHAN, Asharul Islam. Exploring *big data* governance *Frameworks*. **Procedia Computer Science**, v. 141, p. 271-277, 2018.
- ALHASSAN, Ibrahim; SAMMON, David; DALY, Mary. Data governance activities: a comparison between scientific and practice-oriented literature. **Journal of Enterprise Information Management**, v. 31, n. 2, p. 300-316, 2018.
- AL-RUITHE, Majid; BENKHELIFA, Elhadj. A conceptual framework for cloud data governance-driven decision making. In: **2017 International Conference on the Frontiers and Advances in Data Science (FADS)**. IEEE, 2017. p. 1-6.
- ALVES, João. B. da M. **Teoria geral de sistemas**. Preprinted. Florianópolis: [s. n.], 2006. Ebook. Disponível em: https://www.inf.ufpr.br/urban/2019-1_205_e_220/205e220_Ler_Estudar_Discutir/JBosco__Teoria_Geral_de_Sistemas.pdf.pdf.
- ASPINALL, Peter J.; JACOBSON, Bobbie; CASTILLO-SALGADO, Carlos. Establishing and sustaining health observatories serving urbanized populations around the world: scoping study and survey. **The European Journal of Public Health**, v. 26, n. 4, p. 681-686, 2016.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR 15287: informação e documentação: projeto de pesquisa: apresentação**. Rio de Janeiro: ABNT, 2011.
- ATTARD, Judie *et al.* A systematic review of open government data initiatives. **Government Information Quarterly**, v. 32, n. 4, p. 399-418, 2015.
- BARATA, André M. **Governança de dados em organizações brasileiras: uma avaliação comparativa entre os benefícios previstos na literatura e os obtidos pelas organizações**. 2015. Tese (Doutorado em Administração) – Programa de Pós-Graduação em Administração, Universidade de São Paulo, São Paulo, 2015.
- BARBIERI, Carlos. **Uma visão sintética e comentada do Data Management Body of Knowledge (DMBOK)**. Belo Horizonte: Fumsoft, 2013.
- BARBIERI, Carlos. **Governança de dados**: práticas, conceitos e novos caminhos. Alta Books, 2020.

BARBOSA, Jessyca L. P. **Um estudo sobre observatórios através de um mapeamento sistemático da literatura.** 2020. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de Pernambuco, Recife, 2020.

BAYAZIT, Nigan. Investigating design: a review of forty years of design research. **Design Issues**, v. 20, n. 1, p. 16-29, 2004.

BATISTA, Alessandra *et al.* Observatórios de competência. In: **VI Congresso Internacional de Conhecimento e Inovação.** 2016.

BATISTA, Alessandra Duarte *et al.* Observatórios de conhecimento: conceitualização e perspectivas de aplicação na gestão do conhecimento. **Produção em Foco**, v. 7, n. 1, 2017.

BENTO, Patrícia; NETO, Miguel; CÔRTE-REAL, Nadine. How data governance Frameworks can leverage data-driven decision making: a sustainable approach for data governance in organizations. In: **2022 17th Iberian Conference on Information Systems and Technologies (CISTI).** IEEE, 2022. p. 1-5.

BEUTTENMULLER, Gustavo. **Observatórios locais de políticas públicas no Brasil: seu papel na produção, disseminação e transparência das informações.** 2007. Dissertação (Mestrado em Administração Pública e Governo) – Programa de Pós-Graduação em [Nome do Programa], Fundação Getúlio Vargas, São Paulo, 2007.

BRUSEBERG, Anne.; MCDONAGH-PHILP, Deana. Focus groups to support the industrial/product designer: a review based on current literature and designers' feedback. **Applied Ergonomics**, v. 33, p. 27-38, 2002.

ÇAĞDAŞ, Volkan; STUBKJÆR, Erik. Design research for cadastral systems. **Computers, Environment and Urban Systems**, v. 35, n. 1, p. 77-87, 2011.

CAMPOS, P. A. **Dados abertos governamentais: desafios na publicação.** 2018. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2018.

CASAES, Júlio César Costa *et al.* **Governança de dados abertos governamentais: framework conceitual para as universidades federais, baseado em uma visão sistêmica.** 2019.

CHAPPLE, Mike. Speaking the same language: Building a data governance program for institutional impact. **Educause Review**, v. 48, n. 6, p. 14-16, 2013.

DAMA, I. **DAMA-DMBOK: Data management body of knowledge.** **Technics Publications, LLC**, [s. l.], 2017.

DASGUPTA, Avirup; GILL, Asif; HUSSAIN, Farookh. A conceptual framework for data governance in IoT-enabled digital IS ecosystems. In: **8th International Conference on Data Science, Technology and Applications**. SCITEPRESS–Science and Technology Publications, 2019.

DE HAES, Steven; VAN GREMBERGEN, Wim. An exploratory study into IT governance implementations and its impact on business/IT alignment. **Information Systems Management**, v. 26, n. 2, p. 123-137, 2009.

DE LA VEGA, Iván. Un observatorio de ciencia, tecnología e innovación para venezuela. **Cuadernos del CENDES**, v. 19, n. 51, p. 65-81, 2002.

DEMARQUET, Gilles. Five key reasons enterprise data governance matters to finance and seven best practices to get you there. **Journal of Corporate Accounting & Finance**, v. 27, n. 2, p. 47-51, 2016.

DE PAIVA GONÇALVES, Andréa; GASPAR, Marcos Antonio; CARDOSO, Marcos Vinícius. Governança de tecnologia da informação: uma análise do nível de maturidade em empresas atuantes no Brasil. **Gestão e Projetos: GEP**, v. 7, n. 1, p. 56-69, 2016.

DGI. **Data Governance Institute**. [S. I.], 2021. Disponível em: <https://datagovernance.com/the-data-governance-basics/definitions-of-data-governance/>. Acesso em: 17 mar. 2024.

DRESCH, Aline. **Design science e design science research como artefatos metodológicos para engenharia de produção**. Dissertação (Mestrado em Engenharia de Produção) - Programa de Pós-Graduação em Engenharia de Produção – UNISINOS, Novo Hamburgo, 2013.

FERNANDES, Aguinaldo Aragon; DINIZ, Jose Luis; ABREU, Vladimir Ferraz de. **Governança Digital 4.0**. Rio de Janeiro: Brasport Livros e Multimídia Ltda, 2019.

FISHER, Tony. **The data governance maturity model**: establishing the people, Policies and Technology That Manage Enterprise Data. 2007.

FU, Xin *et al.* Data governance in predictive toxicology: a review. **Journal of Cheminformatics**, v. 3, p. 1-16, 2011.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2008.

GROVE, Susan K.; BURNS, Nancy; GRAY, Jennifer R. **The practice of nursing research-E-book**: appraisal, synthesis, and generation of evidence. Elsevier Health Sciences, 2012.

GUSMÃO, Regina. Estruturas e dispositivos nacionais de produção e difusão de indicadores de CT&I: deficiências e possíveis avanços. **Parcerias Estratégicas**, Brasília, v. 20, p. 1075–1098, 2005.

HAIDER, Waqar; HAIDER, Abrar. Governance structures for engineering and infrastructure asset management. In: **2013 Proceedings of PICMET'13: Technology Management in the IT-Driven Services (PICMET)**. IEEE, 2013. p. 1229-1238.

HANEEM, Faizura *et al.* Determinants of master data management adoption by local government organizations: An empirical study. **International Journal of Information Management**, v. 45, p. 25-43, 2019.

HEVNER, Alan *et al.* Design science research in information systems. **Design Research in Information Systems: Theory and Practice**, p. 9-22, 2010.

HUFF, Elena; LEE, John. Data as a strategic asset: Improving results through a systematic data governance framework. In: **SPE Latin America and Caribbean Petroleum Engineering Conference**. SPE, 2020. p. D031S013R001.

JANSSEN, Marijn *et al.* Data governance: organizing data for trustworthy artificial intelligence. **Government Information Quarterly**, v. 37, n. 3, p. 101493, 2020.

JONES, David T. Data governance framework strategic plan. **Philadelphia: Department of Behavior Health and Intellectual disAbility Services**, Dec, 2018.

KARISMA, Karisma; TEHRANI, Pardis Moslemzadeh. Data protection governance framework: A silver bullet for blockchain-enabled applications. **Procedia Computer Science**, v. 218, p. 2480-2493, 2023.

KHATRI, Vijay; BROWN, Carol V. Designing data governance. **Communications of the ACM**, v. 53, n. 1, p. 148-152, 2010.

KIM, Seok-soo. A Case study of implementation data governance for enterprise architecture. **Journal of Information Technology and Architecture**, v. 8, n. 3, p. 255-265, 2011.

KIM, Hee Young; CHO, June-Suh. Data governance framework for *big data* implementation with a case of Korea. In: **2017 IEEE International Congress on Big data (BigData Congress)**. IEEE, 2017. p. 384-391.

KIVUNJA, Charles. Distinguishing between theory, theoretical framework, and conceptual framework: a systematic review of lessons from the field. **International Journal of Higher Education**, v. 7, n. 6, p. 44-53, 2018.

KUMAR, Sand. Data governance: an approach to effective data management. **White Paper, Satyam Computer Services, Ltd**, 2008.

KUSUMASARI, Tien Fabrianti *et al.* Design guidelines and process of metadata management based on data management body of knowledge. In: **2021 7th International Conference on Information Management (ICIM)**. IEEE, 2021. p. 87-91.

LACERDA, Daniel Pacheco *et al.* Design science research: método de pesquisa para a engenharia de produção. **Gestão & Produção**, v. 20, p. 741-761, 2013.

- LADLEY, J. 2019. **Data governance**: how to design, deploy, and sustain an effective data governance program. Elsevier Science.
- LÓPEZ, Juan Ignacio Torregrosa et al. **Proyecto de creación de un observatorio ambiental de la industria de bolivia**. 2012.
- MACEDO, Marcelo; DE SOUZA, Mariane Ramos. Teoria, modelos e *frameworks*: conceitos e diferenças. In: **Anais do Congresso Internacional de Conhecimento e Inovação—CIKI**. 2022.
- MAIZLISH, Bryan; HANDLER, Robert. **IT (information technology) portfolio management step-by-step**: Unlocking the business value of technology. John Wiley & Sons, 2005.
- MANSON, Neil J. Is operations research really research?. **Orion**, v. 22, n. 2, p. 155-180, 2006.
- MARCH, Salvatore T.; SMITH, Gerald F. Design and natural science research on information technology. **Decision Support Systems**, v. 15, n. 4, p. 251-266, 1995.
- MARCONI, Marina A.; LAKATOS, Eva. M. **Metodologia científica**. São Paulo: Atlas, 2007.
- MARTINS, Jorge. Algumas questões em torno da problemática dos Observatórios. v. 11, p. 2017, 2007.
- MASHOUFI, Mehrnaz et al. Data quality in health care: main concepts and assessment methodologies. **Methods of Information in Medicine**, v. 62, n. 01/02, p. 005-018, 2023.
- MATOS, João Filipe; PEDRO, Neuza. De que falamos quando falamos de framework na investigação em educação (matemática). 2008
- MIKALEF, Patrick et al. The role of information governance in *big data* analytics driven innovation. **Information & Management**, v. 57, n. 7, p. 103361, 2020.
- MINAYO, Maria Cecilia de Souza. **O desafio do conhecimento**: metodologia de pesquisa social (qualitativa) em saúde. 1989.
- MORA, José Ferrater. **Dicionário de filosofia**. 4 [S.I.]: Edições Loyola, 2001.
- MOREIRA, Fernanda; FREIRE, Patrícia de Sá; AGUIAR, Ranieri. Digital governance strategy: the relationship between knowledge and federal public governance. In: **INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS & TECHNOLOGY**, 14., 2017. **Proceedings [...]**, n. 130, p. 1165-1182
- NEWMAN, David; LOGAN, Debra. Gartner introduces the EIM maturity model. **Gartner Research Publication**, ID, n. G00160425, 2008.

NURPULAE LA, Lela *et al.* Evaluation of IT governance to support IT operation excellent based on COBIT 4.1 at the PT Timah Tbk. In: **2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)**. IEEE, 2016. p. 336-339.

OPEN KNOWLEDGE FOUNDATION. Disponível em: <https://okfn.org/en/>. Acesso em: 10 jun, 2023.

OPPENHEIM, Charles; STENSON, Joan; WILSON, Richard MS. Studies on information as an asset I: definitions. **Journal of Information Science**, v. 29, n. 3, p. 159-166, 2003.

ORTEGA, Cristina; DEL VALLE, Roberto San Salvador. Nuevos retos de los observatorios culturales1. **Boletín Gestión Cultural Nº19: Observatorios culturales en el mundo**, v. 1697, p. 073X, 2010.

OTTO, Boris. Managing the business benefits of product data management: the case of Festo. **Journal of Enterprise Information Management**, v. 25, n. 3, p. 272-297, 2012.

OTTO, Boris. Organizing data governance: findings from the telecommunications industry and consequences for large service providers. **Communications of the Association for Information Systems**, v. 29, n. 1, p. 3, 2011.

PORTAL BRASILEIRO DE DADOS ABERTOS. Disponível em: <https://dados.gov.br/home>. Acesso em: 12 jun, 2023.

PEROVANO, Dalton Gean. **Manual de metodologia da pesquisa científica**. Curitiba: Intersaber, 2016.

PHÉLAN, Mauricio. La red observatorios locales de Barcelona, España. Un estudio de casos para diseñar una propuesta nacional. Fermentum. **Revista Venezolana de Sociología y Antropología**, v. 17, n. 48, p. 96-122, 2007.

PLOTKIN, David. **Data stewardship**: an actionable guide to effective data management and data governance. Academic Press, 2020.

POSSAMAI, Ana Júlia. **Dados abertos no governo federal brasileiro**: desafios de transparência e interoperabilidade. 2016. Tese (Doutorado em Ciências Políticas) – Curso de Ciências Políticas – Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, 2016.

REYES, Yoel Ramón Sarmiento; FERNÁNDEZ, Mercedes Delgado; ABREU, Marta Beatriz Infante. Observatorios: clasificación y concepción en el contexto iberoamericano. **Revista Cubana de Información en Ciencias de la Salud (ACIMED)**, v. 30, n. 2, 2019.

SANTORO, Paula; XAVIER, Iara. **Observatórios e sistemas de informação em São Paulo, Brasil**: resultados. São Paulo: Pólis, 2009.

SANTOS, Ana Cristina Gomes. Dados abertos e ciência aberta: como as Universidades Federais brasileiras se apresentam nesse horizonte. Biblos. **Revista de Bibliotecología y Ciencias de la Información**, Pittsburgh, 2020, n. 78, p. 01-16. 2020.

SCHNEIDER, Sérgio; SCHIMITT, Cláudia Job. O uso do método comparativo nas Ciências Sociais. **Cadernos de Sociologia, Porto Alegre**, v. 9, n. 1, p. 49-87, 1998.

SILVA, Antonio Waldmir Leopoldino da *et al.* Observatórios brasileiros de meio ambiente e sustentabilidade: diagnóstico e análise. In: **XV Congresso Latino-Iberoamericana de Gestão de Tecnologia**. 2013.

SILVA, Rodrigo Teixeira. **Governança de dados**: proposta de um framework conceitual para os órgãos de segurança pública do Distrito Federal. 2023. 228 f. Dissertação (Programa Stricto Sensu em Governança, Tecnologia e Inovação) - Universidade Católica de Brasília, Brasília, 2023.

SIQUEIRA, Carlos Eduardo; CARVALHO, Fernando. The observatory of the Americas as a network in environmental and worker health in the Americas. **Ciência & Saúde Coletiva**, v. 8, p. 897-902, 2003.

SOARES, Sunil. A Framework that focuses on the data in big data governance. **IBM Data Management**, 2012.

SOARES, Lilian Campos; FERNEDA, Edilson; DO PRADO, Hércules Antonio. Observatórios: um levantamento do estado do conhecimento. **Brazilian Journal of Information Science**, v. 12, n. 3, p. 86-110, 2018.

TALLON, Paul P.; RAMIREZ, Ronald V.; SHORT, James E. The information artifact in IT governance: toward a theory of information governance. **Journal of Management Information Systems**, v. 30, n. 3, p. 141-178, 2013.

THOMAS, Gwen. **How to use the DGI data governance framework to configure your program**. The Data Governance Institute, 2009.

TREMBLAY, Monica. C.; HEVNER, Alan R.; BERNDT, Donald J. Focus groups for artifact refinement and evaluation in design research. **Communications of the Association for Information Systems**, v. 26, n. June, p. 599-618, 2010.

VAN AKEN, Joan Ernst. Management research as a design science: articulating the research products of mode 2 knowledge production in management. **British journal of management**, v. 16, n. 1, p. 19-36, 2005.

YUSOFF, Muhamad Saiful Bahri. ABC of content validation and content validity index calculation. **Education in Medicine Journal**, v. 11, n. 2, p. 49-54, 2019.

ZHANG, Qingqiang; SUN, Xinbo; ZHANG, Mingchao. Data matters: A strategic action framework for data governance. **Information & management**, v. 59, n. 4, p. 103642, 2022.

ZORRILLA, Marta; YEBENES, Juan. A reference framework for the implementation of data governance systems for industry 4.0. **Computer Standards & Interfaces**, v. 81, p. 103595, 2022.

WADA, Maria de Fatima Rufino. **A governança de TI como estratégia para fazer melhor uso do sistema integrado TOTVS RM**. 2022. Dissertação (Mestrado em Tecnologias da Inteligência e Design Digital) - Programa de Estudos Pós-Graduados em Tecnologias da Inteligência e Design Digital da Pontifícia Universidade Católica de São Paulo, São Paulo, 2022.

WANG, Richard Y. A product perspective on total data quality management. **Communications of the ACM**, v. 41, n. 2, p. 58-65, 1998.

WANG, Richard Y.; STRONG, Diane M. Beyond accuracy: what data quality means to data consumers. **Journal of Management Information Systems**, v. 12, n. 4, p. 5-33, 1996.

WEILL, Peter; ROSS, Jeanne W. IT governance: how top performers manage IT decision rights for superior results. **Harvard Business Press**, 2004.

WING, Jeannette M. The data life cycle. **Harvard Data Science Review**, v. 1, n. 1, p. 6, 2019.

WOOK, M *et al.* Exploring big data traits and data quality dimensions for big data analytics application using partial least squares structural equation modelling. **J. Big Data**, vol. 8, pp. 1–15, 2021.

APÊNDICE A

Abaixo estão as perguntas realizadas ao gestor para avaliação do modelo conceitual de governança de dados para o Observatório.

1. Você considera a governança de dados importante para o Observatório? Por quê?
2. As dimensões do modelo (qualidade de dados, segurança de dados, stewardship e gestão de metadados) são suficientes para um programa de governança de dados do Observatório?
3. Os fatores influenciadores são suficientes para um programa de governança de dados do Observatório?
4. Qual(is) outro(s) elemento(s) você traria para o modelo? Justifique sua resposta?
5. De maneira geral, o modelo conceitual é aderente ao Observatório?
6. Você considera que este modelo trará benefícios para o Observatório?