

**IMPACTO DA GOVERNANÇA DE DADOS NO USO DE BUSINESS INTELLIGENCE:
Survey com diferentes categorias de usuários e desenvolvedores de uma grande
empresa brasileira de logística**

PAULO AUGUSTO DE SOUZA MOREIRA

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS (PUC MINAS)

RODRIGO BARONI DE CARVALHO

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS (PUC MINAS)

HUMBERTO ELIAS GARCIA LOPES

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS (PUC MINAS)

IMPACTO DA GOVERNANÇA DE DADOS NO USO DE *BUSINESS INTELLIGENCE*: *Survey* com diferentes categorias de usuários e desenvolvedores de uma grande empresa brasileira de logística

1. INTRODUÇÃO

Face ao aumento da capacidade de armazenamento e consequente maior disponibilidade e volume de geração de dados, as organizações têm buscado se tornar empresas mais *data-driven*, ou seja, empresas que tomam decisões operacionais e estratégicas baseadas em dados. De acordo com estudo da *Harvard Business Review*, encomendado pela Google, empresas que tem suas ações baseadas em dados se mostraram mais resilientes e preparadas para manter sua performance durante o período de pandemia do COVID-19, sendo que o uso de dados se provou fundamental para a rápida tomada de decisões em momentos críticos (Bak, 2023). Ainda de acordo com este estudo, organizações líderes de mercado estão usando dados e Inteligência Artificial (IA) para potencializar todos os demais objetivos estratégicos, apresentando reflexos na melhoria da eficiência operacional, receita, retenção de clientes, satisfação dos empregados e custos (Bak, 2023).

As tecnologias de *Business Intelligence* (BI) têm sido utilizadas nas organizações devido à necessidade de tomar decisões de forma mais rápida e assertiva, estendendo seu uso das informações estratégicas para a execução de tarefas operacionais, aumentando assim o número de profissionais que necessitam usá-las (Imhoff e White, 2011; Böhringer et al., 2010). Aliada às constantes mudanças dos negócios, a demanda por profissionais alocados nas equipes de Tecnologia da Informação (TI), mais especificamente por especialistas em BI, tem aumentado, gerando um gargalo na produtividade das organizações (Kobielus, 2009).

Por outro lado, de acordo com um estudo feito pela Mckinsey, a média das grandes organizações gastam 29% do tempo de seus empregados em tarefas que não geram valor devido à baixa disponibilidade e qualidade de dados (Petzold et al., 2020). Uma pesquisa feita pela Enterprise Strategy Group (2022) com 220 grandes empresas que possuem mais de 1.000 empregados e faturamento anual maior que US\$ 100 MM e que adotaram uma governança de dados mostra que 41% dessas empresas adotaram a governança de dados para a melhoria da qualidade dos dados e 25% para suportar uma melhor tomada de decisão, fatores que favorecem uma utilização mais adequada do *Self-Service Business Intelligence* (Leone e Walker, 2022).

A partir disso, as tecnologias de *Self-Service Business Intelligence* (SSBI) surgem, permitindo aos usuários da organização acessarem os dados e conduzirem suas próprias análises sem a necessidade de envolvimento dos departamentos de TI (Lennerholt et al., 2018), promovendo um empoderamento dos usuários para a exploração dos dados (Alpar e Schulz, 2015). As tecnologias de SSBI visam remover gargalos devido ao aumento dos dados disponíveis nas organizações e o aumento dos usuários, e desde então passaram a ser consideradas como um movimento no campo do BI (Lennerholt et al., 2023).

Apesar dos benefícios previstos nas tecnologias de SSBI, sua implantação apresenta vários desafios. Alpar e Schulz (2015) e Lennerholt et al. (2018) argumentam ser necessário ter pesquisas mais detalhadas sobre como as organizações devem implementar o SSBI. Lennerholt et al. (2018) apresentam os principais desafios envolvidos, que podem ser agrupados em duas categorias: acesso e uso dos dados, e independência dos usuários. Os desafios relacionados ao acesso e uso dos dados abrangem a necessidade de tornar as fontes de dados fáceis de acessar, com critérios para que o usuário possa selecionar os dados a serem utilizados com base em sua qualidade, das fontes corretas, e com políticas de governança de dados implementadas para garantir a integridade e segurança dos dados. Já os desafios relacionados à independência dos usuários tratam sobre tornar as ferramentas de BI fáceis de utilizar, os resultados fáceis de

consumir e o treinamento dos usuários nas melhores práticas de uso dos dados para gerar as melhores tomadas de decisão.

Entre as possíveis soluções para esses desafios do uso do SSBI, entende-se que a governança de dados pode ser um caminho, visando a disponibilização dos dados com qualidade, de forma correta e garantindo uma correta gestão de acesso aos dados. A implementação de uma governança de dados eficaz pode auxiliar as organizações a habilitar uma melhor tomada de decisão, diminuir o atrito organizacional, proteger as necessidades dos stakeholders, construir padrões, reduzir custos e garantir a transparência dos processos (Al-Ruithe e Benkhelifa, 2017). Além disso, a governança de dados permite que as organizações se adequem e garantam a correta utilização dos dados, com a disponibilização do dado para quem de fato precisa e o controle dos objetivos para os quais os dados estão sendo utilizados, respeitando regulamentações relacionadas ao correto uso dos dados, como no caso brasileiro a Lei Geral de Proteção de Dados (LGPD) (Abraham et al., 2019).

A relação entre a governança de dados e o SSBI é um assunto com alto interesse na indústria, porém atualmente faltam pesquisas acadêmicas sobre esse assunto (Clarke et al., 2016). Clarke et al. (2016) apresentam um *framework* que relaciona o SSBI e a governança dos consumidores de dados, principalmente focado no controle do treinamento das pessoas que farão uso do dado para então terem acesso ao mesmo.

Gröger (2018) apresenta ainda a necessidade do estudo da aplicação da governança para *analytics* e BI, e a necessidade da construção de um *framework* para implantação que apresente alguns aspectos relevantes: balancear a acurácia dos dados e a flexibilidade do uso, integrar o modelo às plataformas de BI já existentes na organização e gerenciar o processo de tomada de decisão baseada nos dados.

Dessa forma, torna-se necessária a realização de estudos aprofundados sobre o impacto da implantação da governança de dados nas organizações e a utilização de ferramentas de SSBI. Lennerholt et al. (2018) e Stodder (2015) abordam que uma governança de dados eficaz pode fazer com que os usuários entendam o valor dos dados, evitando a criação de estruturas de *shadow IT*, que podem gerar problemas para a organização. *Shadow IT* é um cenário comumente encontrado nas organizações onde hardwares, softwares ou serviços de tecnologia são utilizados no ambiente de trabalho sem a explícita aprovação ou conhecimento dos responsáveis na organização (Haag e Eckhardt, 2017), caracterizando uma estrutura paralela, redundante ou até mesmo conflitante com a estrutura oficialmente homologada na organização.

Considerando os aspectos e critérios apresentados sobre a governança de dados e os fatores motivadores para a implantação de ferramentas de SSBI nas organizações, esta pesquisa tem o objetivo de analisar o impacto da governança de dados no uso de ferramentas de BI confrontando a avaliação de usuários desenvolvedores e consumidores de SSBI embarcados em um modelo de governança estabelecido na organização, usuários desenvolvedores e consumidores de SSBI não embarcados no modelo de governança e consumidores de relatórios de BI que não fazem uso de ferramentas *self-service*, a fim de comparar os benefícios e restrições percebidas pelos mesmos que afetam a tomada de decisão na organização.

Para além desta introdução, o artigo está dividido da seguinte forma: a seção 2 compreende a fundamentação teórica sobre os temas de *Business Intelligence* e Governança da Dados; a seção 3 explica o método de pesquisa que foi aplicado; a seção 4 apresenta a análise dos resultados obtidos no *survey*; na seção 5 é feita a discussão dos resultados obtidos; e na seção 6 são apresentadas as contribuições teóricas e práticas da pesquisa.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. Business Intelligence

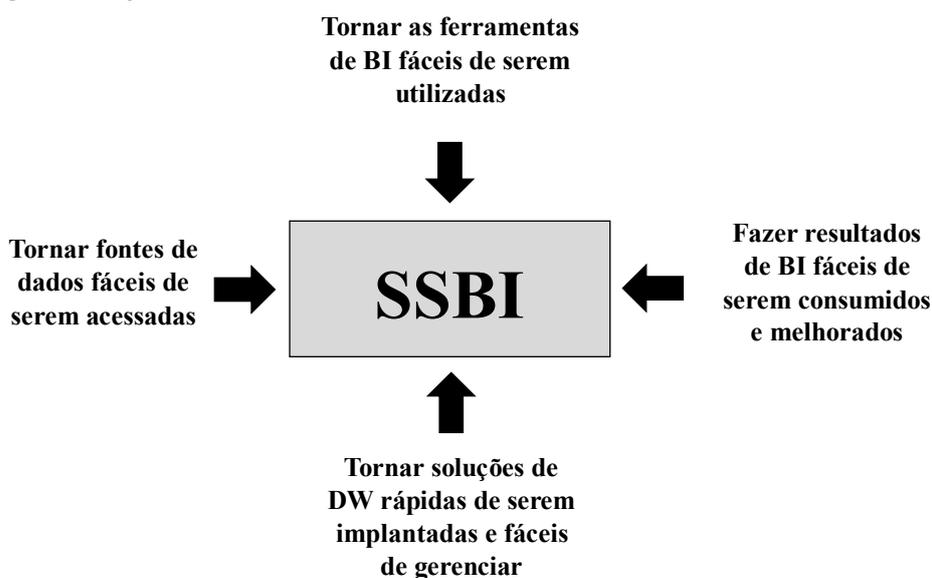
O *Business Intelligence* (BI) compreende o conjunto de ferramentas utilizadas desde o processo de ETL (*Extract, Treat and Load*), que extraem os dados nos sistemas de origem, tratam os dados e carregam em um repositório, passando pelas ferramentas de *Data Warehouse*, que respondem pelo armazenamento dos dados tratados, ferramentas de *Data Mart*, onde os dados são separados por assuntos de interesse, e por último, as ferramentas de gestão da informação, com a construção de relatórios analíticos e análises *ad-hoc* (Ereth e Kemper, 2016).

Um conceito posterior ao BI é o *Business Analytics* (BA), que por sua vez trata-se de metodologias e ferramentas que utilizam estatística e análises quantitativas, modelos exploratórios e preditivos para suportar a tomada de decisão (Cao et al., 2015). Assim, BA, diferente do BI, utiliza-se de ferramentas para armazenamento de grandes volumes de dados (*big data*), passando por distribuição dos dados em bancos de dados não relacionais (NoSQL) e de grafos, utilizados quando bancos de dados relacionais não atendem à finalidade, e por fim, ferramentas de análises preditivas e prescritivas de dados.

Assim, pode-se definir *Business Intelligence e Analytics* (BI&A) como o conjunto de tecnologias, práticas e metodologias com uso de ferramentas estatísticas e análise de dados para apresentação de informações que suportam análises de negócio, mercado e uma melhor tomada de decisão (Chen et al. 2012).

Com a crescente utilização de ferramentas de BI&A e com a baixa necessidade de conhecimentos técnicos e estatísticos para a construção de análises de BI, surgiram ferramentas que implementaram o conceito de *Self-service BI* (SSBI), principalmente vinculadas à etapa de gestão da informação. A principal ideia do SSBI é o empoderamento de usuários casuais, também referenciados como *Citizen Developers*, para criação dos próprios relatórios em sistemas de BI de fácil utilização, sem a necessidade de longas tratativas com os departamentos de TI (Michalczyk et al., 2020). A tecnologia de SSBI se apoia no conceito tecnológico de “faça você mesmo” (do inglês, *Do It Yourself* – DIY) e apresenta 4 objetivos principais a serem atingidos com sua utilização (Imhoff e White, 2011) ilustrados na Figura 1.

Figura 1 - Objetivos do SSBI



Fonte: Imhoff, C., White, C. (2011).

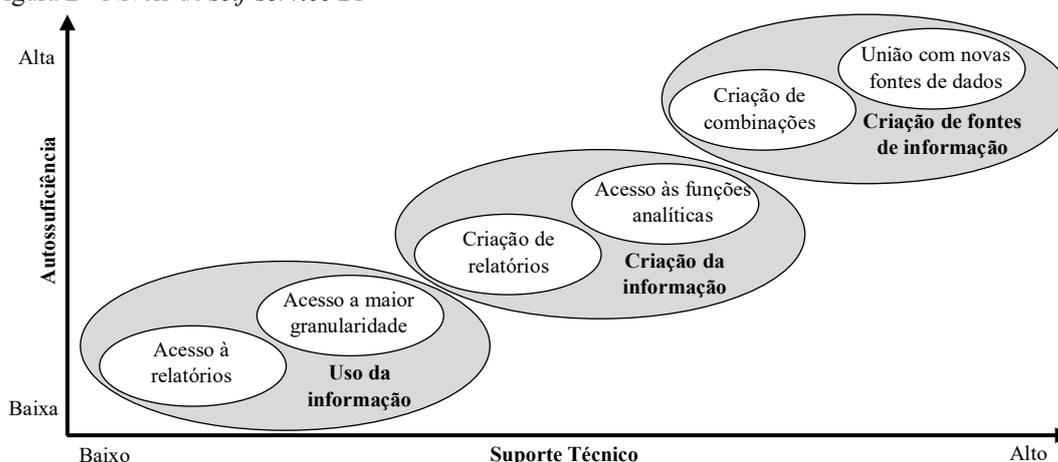
Assim, com base no entendimento de Imhoff e White (2011), pode-se estabelecer hipóteses referentes aos objetivos do uso da tecnologia de SSBI:

- H1: A utilização de ferramentas de SSBI melhora a experiência dos usuários com relação ao consumo dos relatórios;
- H2: A utilização de ferramentas de SSBI melhora a experiência dos usuários com relação à facilidade de uso das ferramentas;
- H3: A utilização de ferramentas de SSBI melhora a experiência dos usuários com relação à velocidade de implementação de relatórios;
- H4: A utilização de ferramentas de SSBI melhora a experiência dos usuários com relação ao acesso às fontes de dados.

Michalczyk et al. (2020) trazem ainda três perspectivas que a literatura aborda na discussão de SSBI: visão centrada nos artefatos, visão centrada no usuário e visão centrada na governança. A visão centrada nos artefatos se refere a uma visão com maior foco nas fontes geradoras dos dados. O objetivo é a interação contínua dos usuários com as aplicações para a busca e integração dos dados situacionais, sem a necessidade de atuação dos programadores (Abelló et al., 2013). Na visão centrada nos usuários, os autores sugerem uma perspectiva com maior foco na facilitação da usabilidade e construção de ferramentas de SSBI para os usuários, a fim de aumentar a aceitação dos sistemas de SSBI (Michalczyk et al., 2020). Por último, a visão centrada na governança apresenta uma perspectiva com maior enfoque na governança e na segurança dos dados, sem tirar a autonomia e velocidade sugerida pelo SSBI. Para tal, ao invés da utilização de *Data Lakes*, que são repositórios centralizados para armazenamento de dados não tratados, em seu formato original (Nambiar e Mundra, 2022), devem ser utilizados, que são repositórios centralizados que armazenam os dados tratados, incluindo vínculo entre diversas fontes, de forma bem estruturada e prontos para o consumo (Chandra e Gupta, 2018).

O conceito de *self-service* pode ser implementado em diferentes etapas, que se dispõem nos níveis de *self-service* para uso dos dados nas organizações: acesso a relatórios preparados ou fontes de dados, acesso direto aos dados, acesso às funções ou a criação de novas fontes de dados (Alpar e Schulz, 2016).

Figura 2 - Níveis de *self-service BI*



Fonte: Alpar, P., Schulz, M. (2016).

De acordo os níveis de *self-service BI* (Figura 2) propostos por Alpar e Schultz (2016), o uso de *self-service* pode ser aplicado em um nível mais baixo de autossuficiência com o acesso à relatórios, onde os usuários possuem acesso às informações que já existem (relatórios existentes), e acesso a maior granularidade dos dados, onde os usuários podem descer em um

menor nível de granularidade da informação. Em um segundo nível, podem ter acesso a criação de relatórios, onde os usuários possuem acesso aos dados na maior granularidade para a criação de novos relatórios, e acesso às funções analíticas, onde podem acessar ferramentas de *advanced analytics* para análises de predição, por exemplo. E, por último, podem ter acesso a criação de combinações para acessar diferentes funcionalidades de reutilização de componentes já preparados pelos profissionais de TI, e a união de novas fontes de dados, onde podem utilizar novas fontes de dados, unindo-as aos dados de fontes já existentes.

O aumento do nível de *self-service* para utilização dos dados na organização por usuários casuais provê maior autonomia a esses usuários, aumentando a possibilidade de experimentações na organização, porém apresenta alguns riscos. Alpar e Schulz (2016) apresentam um exemplo em que uma organização compilou as informações necessárias para a tomada de decisão provenientes de 26.000 relatórios criados após alguns anos utilizando SSBI em 300 relatórios padrões. O exemplo ilustra que a utilização de funcionalidades analíticas para um alto número de usuários não necessariamente é o cenário desejado para obtenção das melhores informações das melhores formas (Alpar e Schulz, 2016).

Além disso, alguns desafios podem ser encontrados na implantação do SSBI nas organizações (Lennerholt et al., 2018). Primeiramente, há o desafio do acesso e uso dos dados que consiste em tornar as fontes de dados fáceis de acessar e utilizar, identificar critérios de seleção dos dados, usar a busca de dados correta, controlar a integridade, segurança e distribuição dos dados, definir políticas para gerenciamento e governança dos dados e preparar os dados para análises visuais. Além disso, as organizações enfrentam o desafio da independência dos usuários, tornando as ferramentas de BI fáceis de se utilizar, os resultados fáceis de consumir e analisar, dar as ferramentas corretas aos usuários corretos e educar os usuários em como selecionar, interpretar e analisar os dados para a melhor tomada de decisão (Lennerholt et al., 2018).

Dessa forma, a importância da governança de dados e da segurança aumentam à medida que são utilizados dados nas organizações. Idealmente, sistemas centralizados e ferramentas de direitos de acessos podem ser utilizados enquanto estruturas de dados temporárias ou permanentes são criadas. Caso contrário, SSBI pode possibilitar falhas na segurança da informação das organizações (Alpar e Schulz, 2016).

2.2. Governança de Dados

A governança de dados começou a ser pesquisada nos anos 1980, tendo ainda atualmente boa parte das pesquisas existentes focadas em algumas áreas específicas. A primeira tentativa de criação de um *framework* para a governança de dados foi publicada em 2007 por Kristin Wende em uma publicação na Conferência Australiana de Sistemas da Informação (ACIS) (Niemi, 2011).

Mesmo sem uma definição uniforme sobre governança de dados, alguns pontos em comum podem ser vistos nas definições encontradas na literatura (Abraham et al., 2019; Khatri et al., 2010; Baltassis et al., 2019; Petzold et al., 2020; DAMA International, 2017), resultando na proposição de uma nova definição consolidada proposta por este trabalho: a governança de dados se destina à definição de papéis e responsabilidades e a atribuição de autoridade, direitos a tomada de decisão e responsabilização sobre a gestão e uso de dados, por meio do estabelecimento de procedimentos, normas e políticas.

É importante avaliar a crescente necessidade de controle sobre o armazenamento de dados, identificando quais dados são armazenados, onde são armazenados, como e por quem são utilizados, tendo em vista a implementação de requerimentos regulatórios, no caso brasileiro pela Lei Geral de Proteção de Dados (LGPD). Esse fator é agravado pela maior

adoção de relatórios *self-service* pelas organizações, o que amplia a utilização dos dados, mas exige maior controle (Abraham et al., 2019).

Pode-se definir 4 fatores objetivos da implementação de uma governança de dados, obtido de diferentes fontes. O DAMA (2017) apresenta o objetivo de segurança e privacidade dos dados de modo que os dados devem estar disponíveis e corretos apenas aos usuários autorizados, respeitando a confidencialidade de seus *stakeholders* e, as regulamentações governamentais, como é o caso da LGPD no Brasil. O objetivo engloba também a propriedade de negócios que auxiliam as organizações na obtenção de vantagens competitivas, a necessidade de fato do uso do dado para a realização do trabalho e as obrigações contratuais previstas em contratos e acordos. Além disso, DAMA (2017) também apresenta como objetivo a necessidade do correto gerenciamento dos metadados, que são definidos como os “dados sobre os dados”, ou seja, descrevem as características e descrições do dado (Khatri e Brown, 2010), e são avaliados pela capacidade de documentar o conhecimento organizacional, integrar metadados de diferentes fontes para o entendimento de suas similaridades e diferenças, a qualidade, segurança e consistência dos mesmos, o acesso para os usuários da organização e a possibilidade de troca de dados entre sistemas.

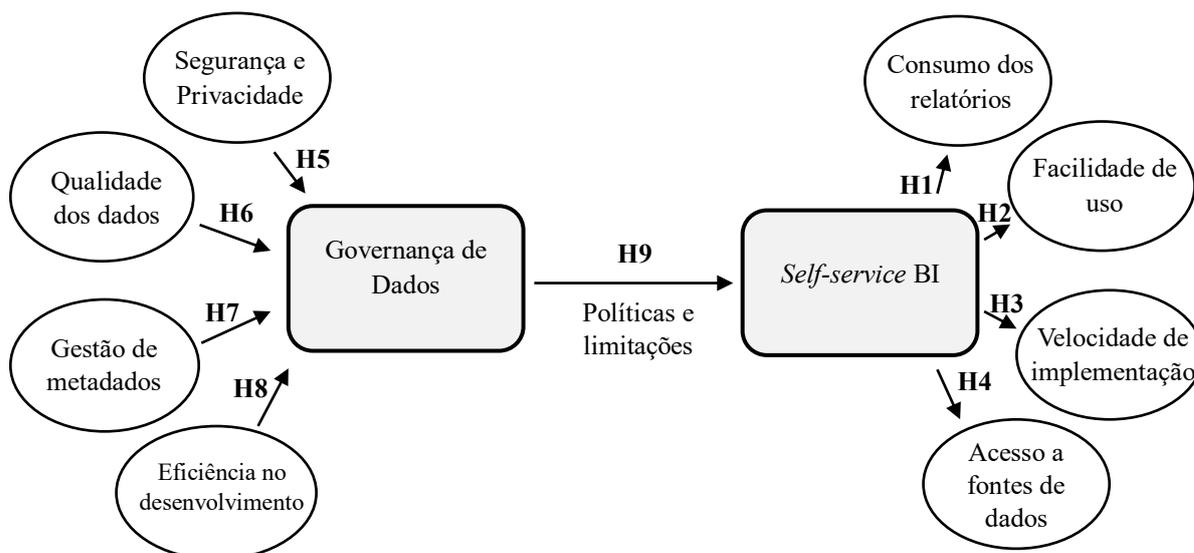
Outro objetivo da governança de dados é a qualidade dos dados, que pode ser avaliada pela sua acurácia, sendo essa a conformidade dos valores dos dados armazenados em relação ao valor real, a pontualidade de atualização do dado quando requisitado, a completude que garante o armazenamento do dado na granularidade necessária e a credibilidade das fontes de dados bem como de seus conteúdos (Khatri e Brown, 2010). E, por fim, o último objetivo se refere à eficiência no desenvolvimento, que pode ser avaliada pelo acesso aos dados, que garante a disponibilidade dos dados frente às necessidades da empresa, salvaguardando a confidencialidade e integridade do dado, e o tratamento do ciclo de vida dos dados. Desta feita, a organização desenvolve formas de otimizar o armazenamento dos dados, mediante aos padrões e necessidades de uso, de modo a reduzir custos.

Pode-se então estabelecer hipóteses referentes aos benefícios da aplicação de um modelo de governança de dados para o uso de ferramentas de SSBI:

- H5: A aplicação da governança de dados impacta positivamente a utilização do SSBI, gerando uma melhor percepção relativa à segurança e privacidade dos dados;
- H6: A aplicação da governança de dados impacta positivamente a utilização do SSBI, gerando uma melhor percepção relativa à qualidade dos dados;
- H7: A aplicação da governança de dados impacta positivamente a utilização do SSBI, gerando uma melhor percepção relativa à gestão de metadados;
- H8: A aplicação da governança de dados impacta positivamente a utilização do SSBI, gerando uma melhor percepção relativa à eficiência no desenvolvimento de produtos de dados;
- H9: A aplicação da governança de dados impacta positivamente a utilização do SSBI, gerando uma melhor percepção geral dos usuários sobre a qualidade das ferramentas.

Assim, o presente estudo visa a avaliação das hipóteses de aplicação de ferramentas de SSBI e o benefício percebido dos usuários referente à governança de dados, conforme modelo conceitual apresentado na Figura 3.

Figura 3 – Modelo Conceitual Preliminar



3. METODOLOGIA

Para avaliação das hipóteses levantadas referentes ao uso das ferramentas de SSBI e a aplicação da governança de dados, foi feita uma pesquisa quantitativa, por meio de *survey*, para realização de um estudo de caso em uma organização brasileira de capital fechado, de grande porte do setor logístico, fruto de um *spin-off* realizado há 13 anos e atualmente possui 7 mil empregados.

A empresa utiliza como ERP (*Enterprise Resource Planning*), desde 2017, o sistema SAP ECC (*Enterprise Core Component*), responsável pela gestão financeira e de processos transacionais administrativos, além de sistemas transacionais desenvolvidos de forma customizada por meio da contratação de fábricas de software. Em janeiro de 2024, os dados armazenados pelos sistemas transacionais totalizavam um armazenamento de 1,02 *petabytes*.

Com base na pesquisa documental, constatou-se que a utilização de produtos de *Business Intelligence* se iniciou em 2017, com a implantação da solução de Microsoft Power BI, e a implantação de um *data warehouse* para melhor captura e tratamento de dados foi feita em 2018. Além disso, a organização implementou em 2019 uma ferramenta de *Data Modeler*, visando o registro dos metadados, para assegurar a correta gestão de conhecimento sobre os desenvolvimentos feitos. Em 2022, tendo em vista a construção de relatórios de SSBI em estruturas *shadow IT*, iniciou-se a migração dos relatórios existentes para um modelo governado, além do desenvolvimento de novos produtos nesse mesmo modelo.

Atualmente, a empresa conta com produtos de *Business Intelligence* tanto desenvolvidos por equipes técnicas de TI quanto desenvolvidos no modelo de SSBI. Com base na avaliação das informações disponíveis em abril de 2024 no Portal do Administrador do Microsoft Fabric e no Portal do Microsoft Azure, ferramentas disponibilizadas pela Microsoft para avaliação e monitoramento do ambiente, a empresa possui atualmente 8.275 relatórios produzidos por tecnologia de SSBI, tanto em estruturas *shadow IT*, quanto seguindo o modelo de governança de dados estabelecido. Desses, 705 relatórios foram utilizados nos últimos 30 dias a contar da data consultada. Além disso, a empresa possui um produto de BI sem a utilização do modelo *self-service*, denominado Centro de Inteligência.

Ainda de acordo com as informações de abril de 2024, a empresa possui 1.139 usuários ativos que utilizam ferramentas de BI. Esses usuários foram separados pela natureza dos produtos que eles consomem, sendo que aqueles usuários que consomem produtos de mais de

uma natureza foram classificados na categoria de produtos que eles mais consomem. Tal critério resultou nos seguintes agrupamentos no universo da pesquisa:

- Usuários de produtos SSBI (*Shadow IT*): 597
- Desenvolvedores de produtos SSBI (*Shadow IT*): 98
- Usuários de produtos SSBI (Governados): 163
- Desenvolvedores de produtos SSBI (Governados): 26
- Usuários do Centro de Inteligência: 255

Pelos conceitos definidos por Michalczyk et al. (2020), o modelo da organização pode ser entendido como um modelo de visão centrada na governança, com o *self-service* no nível 3 – Criação de relatórios, de acordo com o modelo definido por Alpar e Schultz (2016). O *survey* aplicado foi composto por questões que visavam endereçar a percepção de benefício dos construtos de SSBI e governança de dados (Tabela 1) com base nas variáveis de valor geradas.

Tabela 1 – Construtos, variáveis e itens

Construto	Variáveis	Itens	Referências
Self-service BI	Consumo dos Relatórios	Facilidade no consumo	Alpar e Schulz (2016)
		Clareza das definições	Imhoff e White (2011)
		Retroalimentação do conhecimento	Imhoff e White (2011)
	Facilidade de uso	Facilidade de desenvolvimento	Lennerholt et al. (2018)
		Curva de aprendizado	Imhoff e White (2011)
	Velocidade de implementação	Velocidade de implementação	Alpar e Schulz (2016)
		Custo de implementação	Imhoff e White (2011)
	Acesso a fontes de dados	Disponibilidade de acesso	Lennerholt et al. (2018)
Governança de Dados	Segurança e Privacidade	<i>Stakeholders</i>	DAMA (2017)
		Regulamentações	DAMA (2017)
		Propriedade de negócios	DAMA (2017)
		Acesso legitimado	DAMA (2017)
		Obrigações contratuais	DAMA (2017)
	Qualidade dos Dados	Acurácia	Khatri e Brown (2010)
		Pontualidade	Khatri e Brown (2010)
		Completeness	Khatri e Brown (2010)
		Credibilidade	Khatri e Brown (2010)
	Gestão de metadados	Conhecimento organizacional	DAMA (2017)
		Integração entre fontes de dados	DAMA (2017)
		Qualidade dos metadados	DAMA (2017)
		Acesso aos metadados	DAMA (2017)
		Troca de dados	DAMA (2017)
	Eficiência no Desenvolvimento	Acesso aos dados	Khatri e Brown (2010)
Ciclo de vida dos dados		Khatri e Brown (2010)	

Para análise dos resultados do *survey*, primeiramente foi aplicado o método ACP (Análise dos Componentes Principais) de modo a reorientar os dados para se ter poucos componentes que expliquem o maior número possível de informações disponíveis, assim reduzindo as dimensões avaliadas para tornar a visualização dos dados mais direta e a análise mais administrável (Lattin et al., 2011).

Com os componentes definidos, para cada hipótese estudada, foram aplicados testes de correlação entre os componentes gerados pela ACP, variáveis X, e duas variáveis dicotômicas

Y1 e Y2, sendo que a variável Y1 foi avaliada para as hipóteses H1 a H4 e a variável Y2 foi avaliada para as hipóteses H5 a H9, sendo:

- Y1: Os produtos de dados do respondente foram feitos com o uso de ferramentas de SSBI;
- Y2: Os produtos de dados do respondente seguem o modelo de governança de dados da organização.

Para a avaliação das hipóteses H1 a H8, foi calculado Coeficiente de Correlação Ponto Bisserial, por ser adequado para avaliação da correlação entre uma variável contínua X e uma variável dicotômica Y (Ferguson, 1981) e por possibilitar a utilização para uma variável Y não contínua e nem normalmente distribuída (Bunchaft e Kellner, 1999).

Para a avaliação da hipótese H9, foi avaliada a correlação entre os componentes gerados pela ACP das variáveis relacionadas à governança de dados, sendo essas segurança e privacidade, qualidade dos dados, gestão de metadados e eficiência no desenvolvimento. Visto que não se pode assumir uma correlação teórica direta entre os componentes gerados, foi feita análise de regressão logística multivariada. A análise foi feita considerando apenas as variáveis que apresentaram a existência de uma correlação na análise univariada, ou seja, valor-p menor que 0,05 na avaliação do coeficiente de correlação ponto bisserial.

Devido à natureza probabilística da regressão logística multivariada, o R^2 tradicional não é aplicável, e por isso, foi calculado o pseudo R^2 como coeficiente de determinação para avaliar o ajuste do modelo, conforme apresentado por Tjur (2009). Devido à incerteza quanto à normalidade da distribuição da população, para cálculo da amostra mínima necessária, usou-se a fórmula de Slovin (Satriani et al, 2022), considerando uma margem de erro de 5%, resultando em uma amostra mínima de 297.

4. ANÁLISE DOS DADOS

A pesquisa aplicada foi respondida por 368 pessoas, sendo que 216 respondentes utilizam ou desenvolvem produtos de BI com o modelo de governança aplicado, e 268 utilizam ou desenvolvem em produtos de SSBI. Os respondentes estão distribuídos da seguinte forma, com base na categorização dos usuários apresentada:

- Usuários de produtos SSBI (*Shadow IT*): 113
- Desenvolvedores de produtos SSBI (*Shadow IT*): 39
- Usuários de produtos SSBI (Governados): 96
- Desenvolvedores de produtos SSBI (Governados): 20
- Usuários do Centro de Inteligência: 100

O uso de ferramentas de SSBI foi avaliado junto aos usuários com 7 questões aplicadas endereçando as 4 variáveis apresentadas na Tabela 1. Para todas as variáveis, com base na metodologia, foi aplicada a Análise dos Componentes Principais (ACP), de modo a reduzir os itens avaliados em apenas um componente de avaliação. O componente mais representativo foi utilizado para cálculo do Coeficiente de Correlação Ponto Bisserial, correlacionando com a variável Y1. Os resultados da aplicação da metodologia, bem como a representatividade dos componentes de cada variável podem ser encontrados na Tabela 2, conforme aplicados no *software R*.

Tabela 2 – Resultados da avaliação de SSBI

Variável	Quantidade Componentes	Representatividade Componente Principal	Valor-p	Correlação
Consumo dos relatórios	2	85%	0,6289	-0,2528
Facilidade de uso	2	85%	5,31e-09	0,2983
Velocidade de implementação	2	81%	8,65e-18	0,4276
Acesso a fonte de dados	1	100%	0,2779	-0,0567

O alto valor-p encontrado para as variáveis “consumo dos relatórios” e “acesso a fonte de dados” não permite rejeitar a hipótese nula, o que significa que não é possível assumir que haja correlação entre essas variáveis e o uso de ferramentas de SSBI. Para as variáveis “facilidade de uso” e “velocidade de implementação”, foi encontrado valor-p menor que 0,05, permitindo assumir a existência de correlação entre essas variáveis e o uso de ferramentas de SSBI, sendo para a primeira uma correlação positiva e fraca e para a segunda uma correlação positiva e moderada. Dessa forma, há a confirmação das hipóteses H2 e H3 e a rejeição das hipóteses H1 e H4.

A avaliação dos benefícios da aplicação da governança de dados no uso das ferramentas de SSBI, foi feita pela aplicação de 16 questões, endereçando as 4 variáveis apresentadas na tabela 1. Visto a proposta de avaliação da aplicação da governança de dados para o uso de ferramentas de SSBI, para essa análise foram isoladas apenas as respostas dos usuários e desenvolvedores de relatórios que utilizam ferramentas de SSBI.

Para todas as variáveis foi aplicada a Análise dos Componentes Principais (ACP), de modo a reduzir os itens avaliados em apenas 1 componente de avaliação. O componente mais representativo foi utilizado para cálculo do Coeficiente de Correlação Ponto Bisserial, correlacionando com a variável Y2. Os resultados da aplicação da metodologia, bem como a representatividade dos componentes de cada variável podem ser encontrados na Tabela 3, conforme aplicados no software R.

Tabela 3 – Resultados da avaliação de Governança de Dados

Variável	Quantidade Componentes	Representatividade Componente Principal	Valor-p	Correlação
Segurança e Privacidade	5	71%	0,0002	0,2258
Qualidade dos Dados	4	70%	1,31e-12	0,4155
Gestão de Metadados	5	69%	4,10e-15	0,4552
Eficiência no Desenvolvimento	2	84%	0,0226	0,1393

O valor-p menor que 0,05 encontrado para as 4 variáveis permite assumir a existência de correlação entre essas variáveis e a aplicação da governança de dados, sendo uma correlação positiva e fraca para as variáveis “segurança e privacidade” e “eficiência no desenvolvimento”, e positiva e moderada para as variáveis “qualidade dos dados” e “gestão de metadados”. Dessa forma, há a confirmação das hipóteses H5, H6, H7 e H8.

Visto ainda o valor-p menor que 0,05 encontrado para as 4 variáveis relativas à governança de dados, todas as variáveis podem ser utilizadas para aplicação do modelo de regressão logística multivariada, a fim de avaliar o impacto geral da governança de dados no uso de ferramentas de SSBI. Assim, foi feita a aplicação do modelo para análise da correlação dos componentes obtidos para cada uma das variáveis com a variável Y2. O resultado obtido pode ser observado na Tabela 4, conforme aplicado no software R.

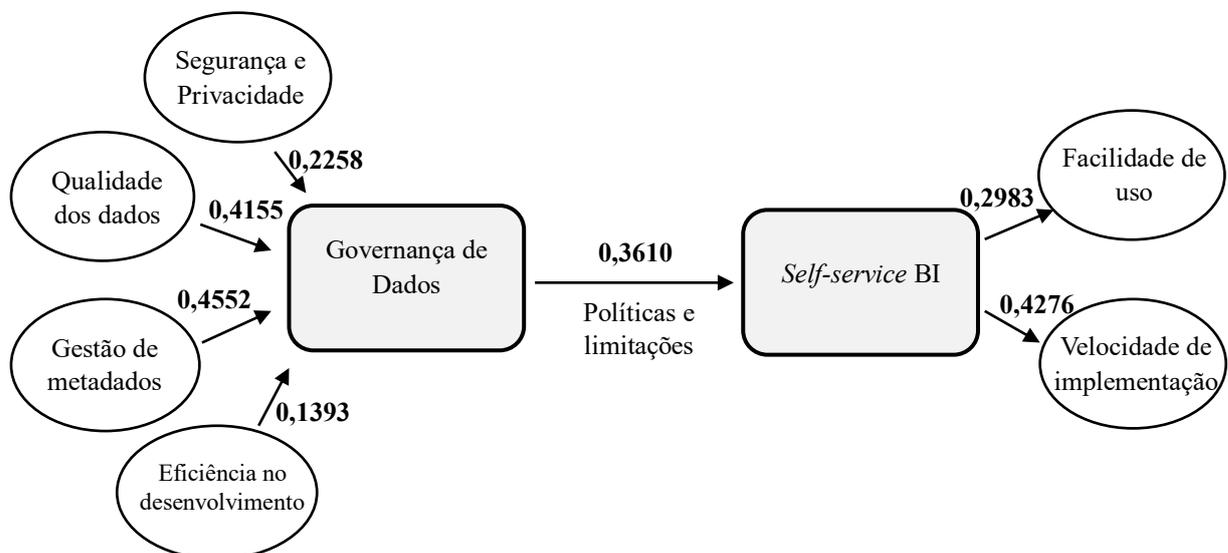
Tabela 4 – Resultados da regressão logística multivariada

Variável	Valor-p
Geral	5,91e-08
Segurança e Privacidade	0,1996
Qualidade dos Dados	0,0048
Gestão de Metadados	6,10e-09
Eficiência no Desenvolvimento	8,42e-06

Conforme resultado obtido, pode-se confirmar uma correlação entre a governança de dados e uma melhor percepção dos usuários de ferramentas de SSBI. O valor obtido para o pseudo R² de Tjur foi de 0,3610, podendo considerar essa correlação positiva e moderada, indicando que a aplicação da governança de dados impacta positivamente o uso de ferramentas de SSBI, confirmando assim a Hipótese H9. Além disso, conforme cálculo de valor-p individual para cada variável dentro do modelo de regressão, pode ser observado um alto valor-p para a variável segurança e privacidade, o que significa que a correlação positiva da aplicação da governança de dados com o uso das ferramentas de SSBI não pode ser explicada por essa variável.

Assim, pode-se construir o modelo final de relação da governança de dados com o uso de ferramentas de SSBI como apresentado na Figura 4.

Figura 4 – Modelo Conceitual Final



5. DISCUSSÃO DOS RESULTADOS

Com base nas avaliações dos resultados obtidos, primeiramente avaliando o uso das ferramentas de SSBI, não foi possível observar a variável “consumo dos relatórios” como um benefício para esse tipo de ferramenta. Isso indica uma igual facilidade e clareza nas definições para consumo das informações tanto em ferramentas de SSBI quanto em ferramentas de BI desenvolvidas pelas equipes de TI.

Outra variável que não apresentou uma correlação com o uso das ferramentas de SSBI foi o “acesso a fonte de dados”. Esse item pode ser observado no caso estudado devido ao estágio de *self-service* em que a empresa se encontra, alocada no nível 3 – Criação de Relatório, com base no modelo definido por Alpar e Schultz (2016). Dessa forma, os *Citizen Developers* possuem menor autonomia para acessar os dados de novas fontes de dados diretamente, estando restritos aos dados fornecidos pelas ferramentas analíticas. À medida que a empresa avança no nível de *self-service*, alcançando o nível 5 – Criação de combinações, pode se esperar uma melhor avaliação nessa variável. Devido às restrições aplicadas pela governança de dados, essa falta de percepção na autonomia para acesso a fontes de dados pelos desenvolvedores e usuários pode levar a um aumento no desenvolvimento de soluções *shadow IT*, por apresentarem uma alternativa mais flexível de acesso mais rápido aos dados, mesmo que com menor qualidade.

Para as outras variáveis foram verificadas correlações positivas, demonstrando que a utilização de ferramentas de SSBI possibilita uma maior facilidade no desenvolvimento de novos relatórios, um aprendizado mais rápido para novos desenvolvimentos, implementação de novos relatórios de forma mais rápida e um menor custo para os desenvolvimentos.

Quando avaliadas as variáveis referentes à aplicação da governança de dados, pode-se observar uma correlação existente e positiva de todas as variáveis com a governança, demonstrando uma percepção de benefício em todas as temáticas quando dessa aplicação. Quando avaliadas individualmente, a variável “eficiência no desenvolvimento” apresenta a correlação mais fraca, o que demonstra que a governança não apresenta grande relevância nessa variável. Conceitualmente, pode haver uma relação entre essa eficiência de desenvolvimento e a baixa percepção relativa ao acesso as fontes de dados observada para as ferramentas de SSBI. Esse fator pode estar relacionado ao histórico ainda recente de uso de ferramentas de BI e da governança de dados na empresa, apresentando baixa maturidade na obtenção de dados quando necessário pelos usuários e pelos desenvolvedores de relatórios na empresa.

Quando avaliada a correlação geral da governança de dados e seus benefícios, pela aplicação da regressão logística multivariada, pode-se observar uma correlação positiva e moderada, demonstrando a percepção positiva de valor da aplicação da governança de dados pelos desenvolvedores e usuários das ferramentas de SSBI. Quando avaliado o cálculo de valor-p individual para cada variável dentro do modelo de regressão, pode ser observado um alto valor-p para a variável segurança e privacidade, o que significa que a correlação positiva da aplicação da governança de dados com o uso das ferramentas de SSBI não pode ser explicada por essa variável. Essa avaliação pôde ser observada na análise univariada quando a segurança e privacidade foi isolada, visto que, apesar de os dados permitirem assumir a correlação entre a variável e o uso de SSBI, esta é considerada fraca.

A variável “segurança e privacidade” possui uma natureza técnica que pode ser pouco percebida pelos usuários de negócio das ferramentas, por ser uma temática específica de TI. Esse benefício pode ser mais bem avaliado caso seja pesquisado com equipes técnicas de governança de dados, governança de TI ou segurança da informação.

Assim, é demonstrada a percepção positiva dos usuários com o uso de ferramentas de SSBI, quando confrontadas com ferramentas tradicionais de BI para uso de equipes de TI, e uma percepção positiva dos benefícios gerados pela aplicação da governança de dados no uso das ferramentas de SSBI, e quais variáveis evidenciam esses benefícios. Durante aplicação do

survey, foi perguntado como os respondentes enxergam a governança de dados na organização. Quando se avalia isoladamente os desenvolvedores de produtos de SSBI governados, observa-se uma percepção positiva do ponto de vista da confiança na qualidade dos dados utilizados. Porém, é percebida uma baixa flexibilidade e autonomia dos produtos, o que conversa com o modelo de *self-service* adotado pela empresa, que apresenta visão centrada na governança. No caso estudado, mostra-se importante evoluir em manter uma segurança e qualidade dos dados, dando maior autonomia para mudanças que competem aos *Citizen Developers*.

Por outro lado, os desenvolvedores de produtos de SSBI não governados possuem pouca visibilidade da governança aplicada na organização, mostrando uma necessidade de melhor comunicação do modelo de governança. Os respondentes também foram questionados das dificuldades enfrentadas para a gestão de indicadores e dados. Os desenvolvedores de produtos de SSBI não governados apontaram principalmente a dificuldade de uso devido à necessidade de alimentação manual dos relatórios, baixa automação, falta de centralização dos dados e muita utilização de dados duplicados, referindo-se à mesma informação, mas sendo buscados de fontes diferentes e com resultados divergentes. Em suma, os fatores apontados são fatores endereçados pela governança de dados e que apresentam maior percepção de valor quando aplicados, pela qualidade dos dados e a gestão dos metadados.

6. CONTRIBUIÇÕES

No aspecto acadêmico, este estudo pode contribuir preenchendo uma lacuna teórica de avaliação da correlação entre a governança de dados e o uso das ferramentas de BI, visto a presença de estudos isolados sobre governança de dados nas empresas e sobre a evolução do uso de ferramentas de BI e SSBI, mas com uma ausência de estudo correlacionando essas duas variáveis. Do ponto de vista de SSBI, este estudo diverge parcialmente do modelo apresentado por Imhoff e White (2011) por não poder confirmar que o SSBI provê maior facilidade no consumo dos resultados e torna as fontes de dados fáceis de serem acessadas. Por outro lado, para a governança de dados, o estudo confirma os objetivos apresentados pelo DAMA (2017) e Khatri et al. (2010), evidenciando os benefícios específicos da governança quando relacionada com o SSBI e reforçando também a percepção pelos usuários e desenvolvedores do valor dos dados, conforme apresentado por Lennerholt et al. (2018) e Stodder (2015).

Quando avaliado o aspecto gerencial, a pesquisa pode contribuir com as empresas que buscam investir na estruturação de uma governança de dados para melhora na utilização das ferramentas de BI. Este estudo possibilita a essas empresas avaliarem os benefícios que podem ser esperados ao fazer a estruturação desse modelo de governança, para justificar o investimento a ser feito, além de demonstrar as dificuldades que podem ser encontradas, bem como os benefícios esperados dentro da evolução no modelo de *self-service* que a empresa pode aplicar.

Como sugestão de pesquisas futuras, o estudo apresentado reforça a importância da governança de dados para endereçar maior segurança e privacidade nos dados, o que não pôde ser observado completamente na pesquisa. Dessa forma, se faz necessária uma pesquisa junto a equipes técnicas a fim de confirmar essa relação da governança com a segurança dos dados. Além disso, a aplicação de um modelo de governança na empresa pode gerar dificuldades e necessidades de restringir as limitações impostas, e se faz necessário um estudo que avalie o processo de migração de um modelo não governado para um modelo de governança de dados aplicada, a fim de avaliar o controle necessário para que essa migração não gere aumento na quantidade de ferramentas de *shadow IT* utilizadas na empresa.

REFERÊNCIAS

- Abelló, A., Darmont, J., Etcheverry, L., Golfarelli, M., Mazón, J. N., Naumann, F., Pedersen, T., Rizzi, S. B., Trujillo, J., Vassiliadis, P., Vossen, G. (2013). Fusion Cubes: Towards Self-service Business Intelligence, *International Journal of Data Warehousing and Mining*, (9:2), pp. 66–88. (DOI: 10.4018/jdwm.2013040104).
- Abraham, R., Brocke, J. V., Schneider, J. (2019). Data Governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, 49, 424-438.
- Alpar, P., Schulz, M. (2015). Self-service Business Intelligence. *Business Information Systems Engineering*. 58. 151-155.
- Al-Ruithe, M., Benkhelifa, E. (2017). Analysis and Classification of Barriers and Critical Success Factors for Implementing a Cloud Data Governance Strategy. *The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017)*. 223-232.
- Bak, J. (2023). *Big on data: Study shows why data-driven companies are more profitable than their peers*. Disponível em: <https://cloud.google.com/blog/transform/data-leaders-more-profitable-innovative-hbr-data>
- Baltassis, E., Gourévitch, A., Quarta, L. (2019). *Good Data Starts with Great Governance*. Disponível em: <https://www.bcg.com/publications/2019/good-data-starts-with-great-governance>
- Böhringer, M., Gluchowski, P., Kurze, C., Schieder, C. (2010). A business intelligence perspective on the future internet. In: *AMCIS 2010 Proceedings*. 1. 267.
- Bunchaft, G., Kellner, S. R. O. (1999). *Estatística sem mistérios*. v.2. Vozes, Petrópolis, RJ, Brasil.
- Cao et al., 2015 – Cao, G., Duan, Y., Li, G. (2015). Linking Business Analytics to Decision Making Effectiveness: A Path Model Analysis. *IEEE Transactions on Engineering Management*, 62(3), 384-395. <https://doi.org/10.1109/TEM.2015.2441875>
- Chandra, P., Gupta, M. K. (2018). Comprehensive survey on data warehousing research. *International Journal of Information Technology*, 10, 217-224.
- Chen, H., Chiang, R. H. L., and Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. *MIS quarterly*, (36:4), pp. 1165–1188. (DOI: 10.2307/41703503).
- Clarke, P., Tyrrell, G., Nagle, T. (2016). Governing self-service analytics. *Journal of Decision Systems*, 25(1), 145-159.
- DAMA International. (2017). *DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition)*. Technics Publications, LLC, Denville, NJ, USA.
- Ereth, J., Kemper, H.G. (2016). Business Analytics und Business Intelligence. *Controlling*. 28. 458-464. (DOI: 10.15358/0935-0381-2016-8-9-458).
- Ferguson, G. A. (1981). *Statistical analysis in psychology and education*. McGraw-Hill Kogakusha, Tokyo.
- Gröger, C. (2018). Building an Industry 4.0 Analytics Platform: Practical Challenges, Approaches and Future Research. *Datenbank-Spektrum*, 18(1), 5-14.

- Haag, S., Eckhardt, A. (2017). Shadow IT. *Business & Information Systems Engineering*, 59(6), 469-473.
- Imhoff, C., White, C. (2011). Self-service Business Intelligence: Empowering Users to Generate Insights. *TDWI Best practices report*. Disponível em: <https://tdwi.org/Research/2011/07/2011YiR-Best-Practices-Report-Q3-Self-Service-Business-Intelligence.aspx?m=1>
- Khatri, V., Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53, 148-152.
- Kobielus, J. (2009). Mighty mashups: do-it-yourself business intelligence for the new economy. *Information and Knowledge Management Professionals*. Disponível em: <https://www.forrester.com/report/Mighty-Mashups-DoItYourself-Business-Intelligence-For-The-New-Economy/RES47806>
- Lattin, J. M., Carroll, J. D., Green, P. E. (2011). *Análise de Dados Multivariados*. Cengage Learning, SP, Brasil.
- Lennerholt, C., Laere, J. V., Söderström, E. (2018). Implementations challenges of self-service business intelligence: a literature review. *51st Hawaii International Conference on System Sciences*. 5055-5063.
- Lennerholt, C., Laere, J. V., Söderström, E. (2023). Success factors for managing the SSBI challenges of the ACQUIRE framework. *Journal of Decision Systems*. 32 (2). 491-512.
- Leone, M., Walker, K. (2022). 2022 State of Data Governance and Empowerment. *ESG Research Insight Paper*. Disponível em: <https://www.erwin.com/docs/2022-esg-state-of-data-governance-and-empowerment-report-analyst-reports-30424.pdf>
- Michalczyk, S., Nadj, M., Azarfar, D., Maedche, A., Gröger, C. (2020). A State-of-The-Art Overview and Future Research Avenues of Self-service Business Intelligence and Analytics. *Twenty-Eighth European Conference on Information Systems (ECIS2020)*. 1-18.
- Nambiar, A., Mundra, D. (2022). An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management. *Big Data and Cognitive Computing*. 6. 132.
- Niemi, E. (2011). *Designing a Data Governance Framework*. Aalto University School of Economics, p. 14. Disponível em: https://www.mn.uio.no/ifi/english/research/news-and-events/events/conferences-and-seminars/iris2013/groups/iris36_submission_4.pdf
- Petzold, B., Roggendorf, M., Rowshankish, K., Sporleder, C. (2020). *Designing data governance that delivers value*. Disponível em: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/designing-data-governance-that-delivers-value>
- SAS Institute. (2018). *The SAS Data Governance Framework: A Blueprint for Success*. Disponível em: <https://www.sas.com/content/dam/SAS/documents/marketing-whitepapers-ebooks/sas-whitepapers/en/sas-data-governance-framework-107325.pdf>
- Satriani, A., Umar, U., Fahlia, F. (2022). The effect of self-organized learning environments (SOLE). Approach on student's leaning motivation and achievement in history lessons. *Prisma Sains: Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram*, 10(3), 776-785.

Stodder, D. (2015). Visual analytics for making smarter decisions faster – applying self-service business intelligence technologies to data-driven objectives. *TDWI Best Practices Report*. Disponível em: https://www.oracle.com/webfolder/s/delivery_production/docs/FY16h1/doc3/TDWI-BPReport-VAnalytics-July2105.pdf

Tallon, P. P., Ramirez, R. V., Short, J. E. (2014). The information artifact in IT governance: toward a theory of information governance. *Journal of Management Information Systems*, 30(3), pp. 141-177.

Tjur, T. (2009). Coefficients of Determination in Logistic Regression Models – A New Proposal: The Coefficient of Discrimination. *The American Statistician*, 63. 366-372.