

**CESAR SCHOOL**

**FERNANDO GONÇALVES DE OLIVEIRA**

**ANÁLISE DO IMPACTO DAS BOAS PRÁTICAS DE GOVERNANÇA  
DE DADOS ALIADO A BIG DATA NA ADMINISTRAÇÃO PÚBLICA  
FEDERAL**

**RECIFE**

**2020**

FERNANDO GONÇALVES OLIVEIRA

**ANÁLISE DO IMPACTO DAS BOAS PRÁTICAS DE GOVERNANÇA  
DE DADOS ALIADO A BIG DATA NA ADMINISTRAÇÃO PÚBLICA  
FEDERAL**

Dissertação apresentada ao programa de Mestrado em Engenharia de Software do Centro de Estudos e Sistemas Avançados do Recife – CESAR SCHOOL, como requisito para a obtenção do título de Mestre em Engenharia de Software.

Orientação: Professor PhD, Felipe Silva Ferraz

**RECIFE**

**2020**

Dissertação de Mestrado apresentada por Fernando Gonçalves de Oliveira ao Programa de Mestrado em Engenharia de Software do Centro de Estudos e Sistemas Avançados do Recife – CESAR SCHOOL, intitulada ANÁLISE DO IMPACTO DAS BOAS PRÁTICAS DE GOVERNANÇA DE DADOS ALIADO A BIG DATA NA ADMINISTRAÇÃO PÚBLICA FEDERAL, orientada por: Felipe Silva Ferraz. A dissertação foi ( ) pela Banca Avaliadora formada por: ( ) e ( ).

Felipe Silva Ferraz  
Orientador

Avaliador(a) Interno(a)

Avaliador(a) Externo(a)

Visto e permitida a impressão

Recife/PE, de de 2020

Felipe Santa Furtado Soares

Coordenador do Mestrado Profissional em Engenharia de Software  
CESAR – Centro de Estudos e Sistemas Avançados do Recife

## **Dedicatória**

A minha mãe Oscalina e irmão Júlio César

À minha esposa Marta Joelma.

Ao meu filho e melhor amigo Kaio Felipe.

Aos meus amigos, colegas, professores e familiares por apoiar minhas decisões e por terem me dado condições para alcançar esse objetivo.

## **Agradecimentos**

Agradeço a Deus.

Agradeço aos professores pelo conhecimento e trocas de ideias, e tornou esse trabalho melhor e em especial ao meu orientador Professor PhD, Felipe Silva Ferraz.

A minha mãe, pelo amor e carinho incondicional.

À minha esposa Marta Joelma, obrigado pelo carinho, pela compreensão e pelo amor que dedica a mim. Eu te amo.

Ao meu filho Kaio Felipe pela inspiração diária de simplicidade e por me lembrar a todo o momento do que vale a pena.

Ao Cesar School que me possibilitou estudar em um ambiente de excelência tecnológico e entidade de referência de ensino superior no Brasil.

A Fábrica de Software Infinity do curso de Mestrado de Engenharia de Software do Cesar School e todos os colegas do curso.

Agradecimentos especiais a José Alberto Sousa Torres, Márcio Fernandes, Rodrigo Hagstrom e Washington Henrique por seu constante apoio e amizade e que deram um empurrãozinho necessário para chegar até aqui.

A todas as pessoas que, direta e indiretamente, me ajudaram na conquista de todos os meus objetivos pessoais e na concretização desta dissertação, o meu sincero agradecimento.

*Fernando Gonçalves de Oliveira*

## Epígrafe

“O que sabemos é uma gota. O que ignoramos é um oceano”.

*Isaac Newton*

## Resumo

Com o avanço da tecnologia, as organizações buscam cada vez mais atentar-se à análise de dados com o intuito de explorar fraquezas e riquezas que contribuam com o crescimento diário das mesmas. Assim sendo, existem correlações entre as demandas técnicas da análise de dados e suas disseminações: quase todas as administrações do setor público usam ferramentas simples de TI para análise, mas uma pequena minoria está envolvida em análises avançadas de dados usando novas tecnologias. No entanto, as organizações estão discutindo seu uso. Agências governamentais (principalmente agências alemãs) são pioneiras nessa área. Quanto mais complexa a análise dos dados, maior a satisfação demonstrada. Tais administrações acreditam que já está se convertendo suas descobertas de análise de dados em benefícios demonstráveis e específicos. Outras administrações observaram apenas efeitos diretos, como critérios aprimorados para a tomada de decisões, riscos reduzidos ou serviços individuais em alguns casos. Desta forma, esta dissertação realiza a Análise do Impacto das Boas Práticas de Governança de Dados aliado a Big Data na Administração Pública Federal - APF, baseando-se no princípio da governança de dados. Além disso, explora-se subsídios para responder quais os objetivos e maiores desafios quanto à implantação de Big Data e/ou Ecossistemas de Big Data nas administrações públicas federais, apontando diretrizes (*guidelines*). A metodologia é concentrada em três fases (mapeamento sistemático das aplicações de Big Data, pesquisa survey aplicada em instituições públicas federais e por fim uma análise baseada em grupo focal especializado em Big Data para definição e aplicação de uma *guideline* como proposta para resolução de barreiras apontadas pelo grupo focal. A metodologia se relaciona com o conhecimento da literatura correlata que pode auxiliar no entendimento da questão em epígrafe. Compreende-se que a implantação de Big Data no setor público precisa estar alinhada com a missão institucional do órgão que se propõe a implantar Big Data, mas que os gestores públicos nem sempre percebem suas instituições como aptas a implantarem projetos de Big Data, o que pode, por sua vez, demandar que seja apresentado todo um conjunto de diferentes projetos de Big Data em diferentes esferas e missões institucionais. Finalmente, se percebe que há pouca literatura relacionada com o caso brasileiro, o que pode destacar a importância da pesquisa para esta dissertação.

**Palavras-Chave:** Big Data, Administração Pública Federal, Ecossistema. Survey.

## **Abstract**

*With the advancement of technology, organizations increasingly seek to pay attention to data analysis in order to explore weaknesses and riches that contribute to their daily growth. As such, there are correlations between the technical demands of data analysis and its dissemination: almost all public sector administrations use simple IT tools for analysis, but a small minority are involved in advanced data analysis using new technologies. However, organizations are discussing its use. Government agencies (mainly German agencies) are pioneers in this area. The more complex the analysis of the data, the greater the satisfaction shown. Such administrations believe that their data analysis findings are already converting into demonstrable and specific benefits. Other administrations observed only direct effects, such as improved criteria for decision making, reduced risks or individual services in some cases. Thus, this dissertation Analyzes of the Impact of Good Data Governance Practices combined with Big Data in the Federal Public Administration - APF, based on the principle of data governance. In addition, subsidies are explored to answer the objectives and major challenges regarding the implementation of Big Data and / or Big Data Ecosystems in federal public administrations, pointing out guidelines. The methodology is concentrated in three phases (systematic mapping of Big Data applications, survey research applied in federal public institutions and finally an analysis based on a focus group specialized in Big Data for the definition and application of a guideline as a proposal for solving pointed barriers by the focus group. The methodology relates to the knowledge of the related literature that can assist in the understanding of the above question. It is understood that the implementation of Big Data in the public sector needs to be aligned with the institutional mission of the body that proposes to implement Big Data, but that public managers do not always perceive their institutions as able to implement Big Data projects, which may, in turn, demand that a whole set of different Big Data projects be presented in different spheres and institutional missions. Finally, it is noticed that there is little literature related to the Brazilian case, which can highlight the imposition importance of research for this dissertation.*

**Keywords:** *Big Data, Federal Public Administration, Ecosystem. Survey.*



## SUMÁRIO

1. INTRODUÇÃO.....	15
1.1. Motivação.....	16
1.1.1 Motivação de mercado e barreiras .....	16
1.1.2 Motivação técnica .....	17
1.2. Problemática e Justificativa.....	18
1.3. Objetivos .....	21
1.3.1 Objetivo Geral.....	21
1.3.2 Objetivos específicos .....	21
1.4. Contribuições .....	21
1.5. Ilustração metodológica .....	22
1.6. Estrutura da dissertação .....	23
2. REFERENCIAL TEÓRICO .....	24
2.1. Governança de dados.....	24
2.1.1 Concepção histórica e sua relevância para o cenário .....	24
2.1.2 Riscos aplicados a governança de dados .....	27
2.1.3 Entendendo sobre o DAMA DMBOK: Conceitos e Diretrizes .....	31
2.1.4 Aplicação da governança de dados.....	33
2.2. Big Data e suas aplicações .....	36
2.2.1 Conceito histórico e atribuições.....	36
2.2.2 Aplicações potenciais de Big Data para a Administração Pública Federal e suas restrições.....	42
2.2.3 Recursos de dados disponíveis para a Administração Pública.....	44
2.2.4 Requisitos da administração pública: Técnicos e não técnicos. ....	46
2.2.5 A importância do Big Data na Administração Pública Federal.....	49
2.3. Governança de dados aliado ao Big Data.....	53
2.3.1 Porque governança de dados é crucial para ambientes de Big Data ....	53
2.3.2 O papel integral da Governança de dados em ambientes de big data ..	56
2.3.3 Como as diferenças de Big Data podem afetar a Governança de dados	59
2.4. Aplicações de governança de dados e Big Data em Instituições públicas federais: Uma visão globalizada .....	61
3. METODOLOGIA E APLICAÇÃO DA PESQUISA .....	65
3.1. Compreendendo as etapas metodológicas .....	65
3.2. Apresentação do Mapeamento Sistemático.....	67

3.3.	Apresentação da Pesquisa Survey .....	68
3.2.1	Etapas da aplicação do questionário .....	72
3.4.	Apresentação da pesquisa Grupo Focal .....	74
4.	RESULTADOS E DISCUSSÕES .....	76
4.1.	Análise da aplicação da pesquisa Survey .....	76
4.2.	Análise da aplicação da pesquisa de Grupo Focal .....	102
4.2.1	Perfil dos participantes.....	102
4.2.2	Avaliação de Prontidão para Big Data – Grupo Focal.....	102
4.2.2	Comentários da Avaliação de Prontidão para Big Data – Grupo Focal ...	105
4.2.3	Entendendo o cenário atual .....	107
4.2.5	Guideline para implantação de Governança de dados e Big Data em Instituições públicas federais .....	108
5.	CONCLUSÕES.....	109
6.	REFERENCIAL BIBLIOGRÁFICO .....	113

## **Lista de imagens**

Figura 1 – Apresentação ilustrativa da metodologia.....	22
Figura 2 - Demonstrativo da utilização de governança de dados.....	25
Figura 3 – Modelo de framework para gestão de dados .....	28
Figura 4 – Modelo DAMA-DMBOK, segundo guia .....	32
Figura 5 - Demonstração de Big Data através dos 5 Vs. ....	38
Figura 6 – Governança de dados integrado com o ambiente corporativo .....	55
Figura 7 – Compreensão das etapas metodológicas. ....	66
Figura 8 – Etapas do mapeamento sistemático. ....	67
Figura 9 – Síntese dos procedimentos de pesquisa bibliográfica.....	68
Figura 10 – A Concepção e a Função do Questionário.....	69
Figura 11 – Etapas da Pesquisa. ....	71

## **Índice de Tabelas**

Tabela 1 - Critérios dos cenários de aplicação para a Administração Pública. ....	42
Tabela 2 – Avaliação de Prontidão para Big Data – Grupo Focal. ....	103

## Lista de Gráficos

Gráfico 1 - Cargo na Administração Pública ocupados pelos entrevistados .....	76
Gráfico 2 – Em qual poder trabalha o entrevistado .....	77
Gráfico 3 - Tempo de experiência com T.I do entrevistado .....	78
Gráfico 4 - Volumetria de armazenamento dos órgãos. ....	80
Gráfico 5 - Projeção de crescimento do banco de dados.....	81
Gráfico 6 - Possui projeto para ambiente próprio de Big Data? .....	82
Gráfico 7 - Diferenciação entre os termos distintos de Big Data. ....	83
Gráfico 8 - Considera que todos os termos acima possuem forte relação com o tema Big Data?.....	84
Gráfico 9 - Tipos de armazenamento de dados do órgão. ....	84
Gráficos 10 – Possui dados não estruturados .....	85
Gráficos 11 - Existe algum problema de negócio que poderia ser resolvido com Big Data? .....	87
Gráficos 12 - Perspectiva para implantação de Big Data nos órgãos. ....	88
Gráficos 13 - O órgão já possui projetos de Big Data? .....	88
Gráficos 14 - O que tem motivado o uso de Big Data no seu órgão público? .....	89
Gráficos 15 - Dificuldades encontradas na implementação do ambiente de Big Data. ....	90
Gráficos 16 - Impactos com a implantação de Big Data.....	91
Gráficos 17 - Os impactos econômicos da adoção de Big Data nos órgãos.....	92
Gráficos 18 - As oportunidades de implementação do Big Data nos órgãos públicos. ....	93
Gráficos 19 - Os impactos que práticas de Big Data podem/poderiam trazer ao desenvolvimento de soluções nos órgãos públicos.....	94
Gráficos 20 - Os impactos citados na questão anterior são mais positivos que negativos nos órgãos públicos? .....	95
Gráficos 21 - A sua resposta à questão anterior envolveu impactos existentes e comprovados (Sim) ou impactos percebidos pelo conhecimento acerca de Big Data conhecido pelo respondente do questionário (Não).....	96
Gráficos 22 - O órgão compartilha conhecimentos e aplicações de Big Data entre todos os setores do órgão? .....	97

Gráficos 23 - A equipe de TI e profissionais de dados compartilha conhecimentos entre todos os setores da organização? .....	97
Gráficos 24 - No ecossistema de Big Data é explorado ou adotado serviços de nuvem para processar ou conduzir o Analytics? .....	98
Gráficos 25 - Quantos profissionais de cientistas de dados o órgão possui servidores públicos? .....	99
Gráficos 26 - Quantos profissionais de cientistas de dados o órgão possui terceirizado? .....	99

**Abreviaturas**

Sigla	Significado
TI	Tecnologia da Informação
TCLE	Termo de Consentimento Livre e Esclarecido

## 1. INTRODUÇÃO

Este trabalho analisa os impactos das boas práticas de Governança de Dados aliado a Big Data na Administração Pública Federal do Brasil, apontando desafios que os órgãos competentes governamentais encontram ao implantar tecnologias que tratam grandes volumes de dados. Tais tecnologias são discutidas ao longo da dissertação com o intuito de desmistificar sua importância e empregabilidades para a concepção das diretrizes sobre Big Data.

Verifica-se na literatura pouca produção relacionada com projetos de Big Data e órgãos governamentais se comparada a pesquisas sobre Big Data e Governança de dados, em setores privados, o que confere relativa relevância e novidade ao tema, embora o setor público esteja adquirindo mais conhecimento sobre o valor potencial derivado do uso da governança de dados e Big Data, à medida que os órgãos componentes de sua estrutura, governos geram e coletam grandes quantidades de dados por meio de suas atividades cotidianas (GONÇALVES & VARELLA, 2018).

No entanto, a análise em grandes volumes de dados, se constitui, sim, em um desafio para os órgãos componentes do setor público, assim como ocorre em organizações privadas. Uma razão para isso é que a legislação acerca de Governo Eletrônico ainda não foi implementada por completo (SOUZA, 2019). Em adição, há muito a ser feito em relação à segurança de TI.

Os benefícios das aplicações relacionadas a Governança de dados e do Big Data no setor público podem ser agrupados em três áreas principais, com base na classificação dos tipos de benefícios (REIS, 2015):

- análise avançada, por meio de algoritmos automatizados;
- melhorias na eficácia, proporcionando maior transparência interna;
- melhorias na eficiência, nas quais melhores serviços podem ser fornecidos com base na personalização dos serviços; e aprender com o desempenho de tais serviços.

Tais apontamentos realizados por Reis (2015) enfatizam a importância da aplicação de tecnologias voltadas para grandes volumes de dados, porque no setor público são tratados vários seguimentos de informações de vários setores diferentes, sendo necessário tal análise para medidas que possam ser beneficiar a curto e a longo prazo.

Algumas ideias iniciais parecem evidentes com base na literatura correlata, como o fato de que desenvolvimentos são necessários nos campos de escalabilidade da análise de dados, descoberta de padrões e aplicativos em tempo real. Ainda, melhorias na procedência do compartilhamento e integração de dados do setor público, e o fornecimento de mecanismos integrados de segurança e privacidade em aplicativos de Big Data.

## **1.1. Motivação**

### **1.1.1 Motivação de mercado e barreiras**

O ambiente de trabalho, foco da pesquisa, é composto de órgãos públicos, ligados ao Governo Federal do Brasil, o qual, por sua vez, é um comprador de produtos e serviços de Tecnologia da Informação - TI na América Latina, sendo um dos clientes.

Assim sendo, o ambiente privado de TI, especialmente na cidade de Brasília, é atingido pelas demandas do Governo Federal do Brasil. E, considerando-as cada vez mais necessárias, as tecnologias de Governança de Dados e Big Data que podem vir a ser aplicadas pelos órgãos de governo, tornam-se necessárias e importantes para auxiliar a busca, dentro deste diapasão, de respostas que permitam uma melhor orientação de esforços de implementação das tecnologias nestes órgãos do Governo Federal, ou mesmo de outras esferas de governo no Brasil.

Algumas características que diferenciam o ambiente de trabalho de órgãos públicos do mercado privado é o fato de não existir concorrência direta entre estes órgãos (o que pode gerar ambiente de cooperação entre os aludidos), pois o setor



público não é um setor produtivo, embora suas despesas representassem uma “fatia” considerável do PIB.

A maior parte da receita do setor é arrecadada por meio de impostos e contribuições sociais. Assim, o impacto das tecnologias de Big Data está relacionado com aspectos de eficiência e eficácia: quanto mais eficiente o setor público, melhores são atendidos os cidadãos, e menos recursos (impostos) são necessários para fornecer o mesmo nível de serviço.

Uma característica particular do ambiente público é que este abrange a aplicação de Governança de Dados e Big Data para fornecer maior transparência, acesso da coletividade a informações de seus governos. Em adição, cidadãos e empresas podem participar das tomadas de decisão por meio de sistemas que forneçam a compilação das informações geradas pelos sistemas de Big Data e Mineração de Dados, gerando opiniões embasadas em dados reais.

Em uma última análise, a qualidade dos serviços prestados, por exemplo, educação, saúde, serviços sociais, políticas ativas e segurança, também pode ser aprimorada com o uso de tecnologias de Governança de Dados e Big Data, com base no aprimoramento da qualidade de direcionamento de recursos e decisões.

### **1.1.2 Motivação técnica**

Projetos de implantação de Governança de Dados e Big Data, assim como qualquer projeto, necessita, estar alinhado com os objetivos estratégicos da organização que o utiliza ou implementa. Tal problemática será abordada ao longo do texto do presente documento. Ou seja, as técnicas de Big Data em si não são suficientes para fornecer subsídios de melhoria da atuação de qualquer organização.

No entanto, aspectos tecnológicos da Governança de Dados e Big Data são abordados, ainda que alguns órgãos não tenham encontrado o perfeito alinhamento entre técnicas e projetos de Big Data e sua missão institucional. Logo, como motivação técnica, pode-se descrever, em princípio, que os benefícios da Governança de Dados e o Big Data no setor público podem ser agrupados em três

áreas principais, com base na classificação dos tipos de benefícios, ou seja, a própria análise do Big Data, os seus processos e melhorias na eficácia e eficiência das rotinas e incrementos nos processos de tomada de decisão.

Para tanto, o uso de sistemas de suporte a tomada de decisão com uso de ferramentas de Big Data pode produzir, por exemplo, conjuntos de resultados que revelam padrões ou mecanismos mais eficientes de previsões de comportamento, o que conduz, desde que trabalhados os resultados de forma correta, em alinhamento com as políticas públicas, melhorar o processo decisório. Tais ferramentas incluem administração de poder computacional, aplicativos de Mineração de Dados, dentre outros.

Nesse contexto, essa área abrange os aplicativos que fornecem, portanto, melhores serviços com melhoria contínua, com base na personalização de serviços e aprendizado com o desempenho desses mesmos serviços.

Tais apontamentos serão mensurados a partir da análise descritiva de um questionário que foi disponibilizado para órgãos públicos evidenciando um conjunto de fatores que podem beneficiar ou restringir as diretrizes de Big Data. Este questionário é parte importante deste objeto de estudo, e foi abordado mais a diante deste trabalho.

## **1.2. Problemática e Justificativa**

O problema foco da pesquisa é responder quais são as restrições e desafios à implantação e uso das tecnologias de Governança de Dados e Big Data no setor público, e conhecer onde os governos podem atuar como catalisadores no desenvolvimento de um ecossistema de dados por meio da abertura de seus próprios conjuntos de dados, gerenciando sua disseminação e uso (objetivos e oportunidades).

O setor público está se tornando cada vez mais consciente do valor potencial a ser obtido com o big data, à medida que os governos geram e coletam grandes quantidades de dados por meio de suas atividades cotidianas (YIU, 2012).

Os benefícios da Governança de Dados e o big data no setor público podem ser agrupados em três áreas principais, com base na classificação dos tipos de benefícios: análise avançada, por meio de algoritmos automatizados; melhorias na eficácia, proporcionando maior transparência interna; melhorias na eficiência, onde melhores serviços podem ser fornecidos com base na personalização dos serviços; e aprender com o desempenho de tais serviços.

Desta forma, examina-se vários fatores e restrições que foram identificados, que podem impulsionar ou interromper o desenvolvimento de big data no setor, dependendo de como eles são abordados. As descobertas, após analisar os requisitos e as tecnologias atualmente disponíveis, mostram que há questões de pesquisa abertas a serem abordadas, ou seja, problemas (como por exemplo, o porque tais tecnologias relacionadas ao big data não foram difundidas pelos órgãos públicos? Ou, o porque o governo negligencia uma base de dados que poderia ser trabalhada a ser favor, com as práticas de big data?, Ou ainda, o porque as empresas que fornecem soluções para esta governança de dados não são acionadas pelos governos, sabendo que internamente não existem projetos que dêem continuidade na governança de dados e conseqüentemente na prática de big data) para o desenvolvimento de tais tecnologias, para que soluções competitivas e eficazes possam ser construídas.

Os principais desenvolvimentos são necessários nos campos de escalabilidade da análise de dados, descoberta de padrões e aplicativos em tempo real. Também são necessárias melhorias na procedência do compartilhamento e integração de dados do setor público. Também é extremamente importante fornecer mecanismos integrados de segurança e privacidade em aplicativos de big data, pois o setor público coleta grandes quantidades de dados confidenciais. Por fim, respeitar a privacidade dos cidadãos é uma obrigação do setor público nacional.

Sabe-se que os governos geram e coletam grandes quantidades de dados por meio de suas atividades cotidianas, como gerenciamento de pensões e pagamentos de subsídios, cobrança de impostos, sistemas nacionais de saúde, registro de dados de tráfego e emissão de documentos oficiais.

Desta forma, o problema leva em consideração as atuais tendências socioeconômicas e tecnológicas, incluindo o aumento da produtividade em um ambiente com restrições orçamentárias significativas, a crescente demanda por serviços médicos e sociais e a padronização e interoperabilidade como requisitos importantes para as aplicações e tecnologias do setor público, buscando salientar o seguinte: **Os órgãos públicos brasileiros estão devidamente preparados para a utilização das diretrizes de governança de dados e big data ?**

Como justificativa, destacam-se as iniciativas de dados abertos, que são um ponto de partida para impulsionar um mercado de dados e das tecnologias de Big Data. Adicionalmente, as necessidades de melhoria de tomada de decisão, acrescida da possibilidade maior participação do cidadão junto aos órgãos de governo.

As políticas ativas na área de dados abertos podem beneficiar o setor privado e facilitando o crescimento desse setor no Brasil. Em contrapartida, isso beneficiará os orçamentos públicos com o melhor aproveitamento da receita tributária de uma crescente indústria de dados na em âmbito nacional.

### **1.3. Objetivos**

#### **1.3.1 Objetivo Geral**

Por meio da análise de boas práticas e aplicações da Governança de Dados e Big Data no setor público brasileiro, em âmbito federal, busca-se delimitar quais os objetivos e maiores desafios e dificuldades relacionadas com a adoção das tecnologias no aludido cenário, através de uma pesquisa do tipo *survey* em instituições coligadas ao tema e através de um grupo focal que ajudará nas aplicações de uma *guideline* como proposta para a adoção de Big Data para estas Instituições.

#### **1.3.2 Objetivos específicos**

Os objetivos específicos consideram um escopo mais focado em aspectos relacionados com as necessidades e com os papéis dos órgãos públicos mediante a sociedade administrada, quais sejam, buscam responder às seguintes questões:

- Qual é o papel da governança de dados no setor público federal?
- Como governança de dados, pode melhorar o cenário problemático e econômico do Brasil?
- Quais são os impactos econômicos das soluções de Big Data no setor público?
- Quais as oportunidades e desafios da adoção de Big Data no setor público?
- Quais os impactos das práticas de Big Data no setor público?
- Evidenciar problemas ligados ao setor público que dificultam as práticas e diretrizes de Big Data;
- Propor uma guideline como diretriz para adoção e prática para Big Data, formada através de uma pesquisa com um grupo focal.

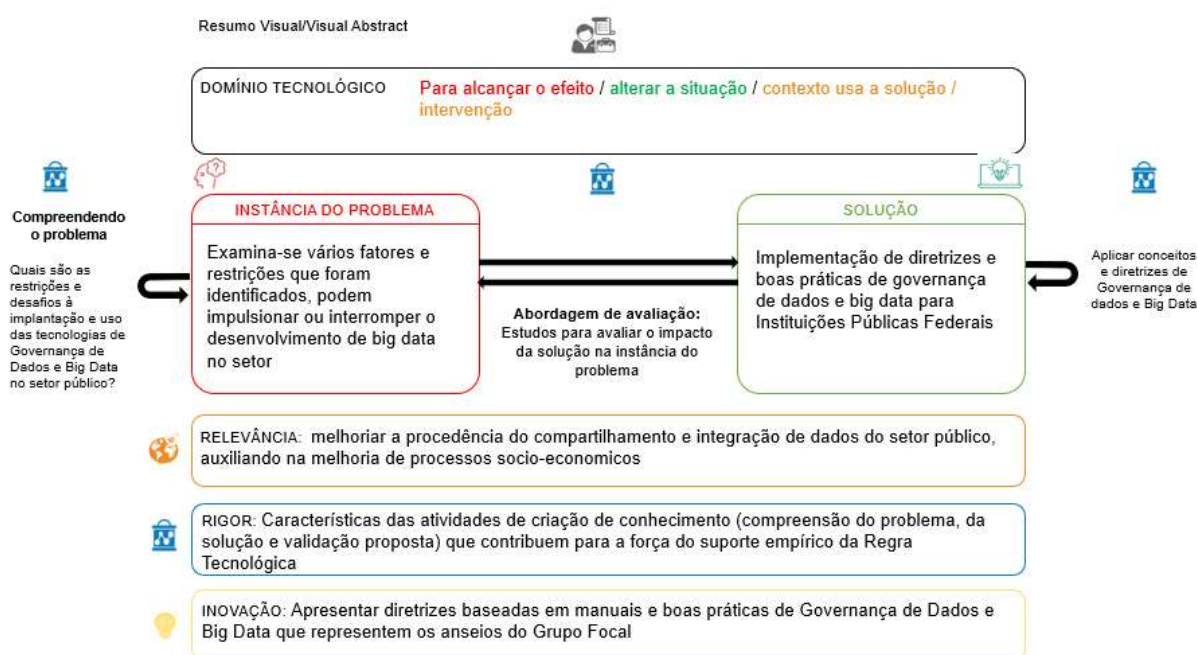
### **1.4. Contribuições**

- Fornecer dados concisos que comprovem que a prática de Big Data, pode interferir de maneira positiva ou não na Administração Pública Federal;

- Contribuir com o estímulo de incremento de eficiência e eficácia dos processos e consequentemente na análise dos dados dos setores públicos;
- Dissertar sobre os processos de Big Data disseminando boas práticas na administração pública, contribuindo com literatura técnica e científica para o corpo acadêmico e futuros leitores que possam se interessar pelo tema.
- Dissertar sobre as diretrizes coligadas a Governança de Dados;
- Evidenciar a importância da governança de dados em conjunto com Big Data;
- Elaborar uma guideline com o propósito de guiar os leitores na adoção de big data em conformidade com governança de dados, aplicadas a instituições públicas federais.

### 1.5. Ilustração metodológica

Figura 1 – Apresentação ilustrativa da metodologia



Fonte: Engstrom et al. (2020) – Modelo adaptado pelo autor

## **1.6. Estrutura da dissertação**

A estrutura desta dissertação pode ser compreendida na seguinte forma: a primeira seção diz respeito a introdução da temática, bem como sua importância e relevância para o meio acadêmico e profissional, relatando a problemática, justificativa, bem como o objetivo central e objetivos específicos que compõe este objeto de estudo. Posteriormente relatam-se as motivações técnicas e de mercado, que impulsionaram o autor para a realização desta dissertação.

Na segunda seção, destina-se ao desenvolvimento do referencial teórico desta dissertação, onde serão apresentados conceitos, diretrizes, aprofundando o conhecimento técnico e prático relacionado à temática deste objeto de estudo.

Na terceira seção, apresenta-se a metodologia que compõe esta dissertação. Compreendem-se suas práticas, métodos e aplicações para a exploração do estudo de caso que será apresentado na quarta seção.

A quarta seção trás as aplicações da metodologia, baseadas em 3 fases, onde: explora-se a pesquisa survey através de um questionário aplicado a instituições publicas federais. Relata-se um questionário aplicado através de um grupo focal e apresenta-se diretrizes impostas como solução da aplicação das tecnologias para Governança de dados e Big Data.

Na quinta seção, destina-se aos resultados e conclusões que esta dissertação apresenta como contribuição dos estudos dirigidos à aplicabilidade do tema na Administração Pública Federal, e logo em seguida apresentam-se as referências bibliográficas que compõe esta dissertação.

## **2. REFERENCIAL TEÓRICO**

Nesta seção são apresentados os tópicos correlatos ao tema tratado, por meio da revisão da literatura. Inicialmente, os tópicos são abordados de forma mais direta, considerando não apenas os conceitos de destinados ao tópico, sua importância no cenário proposto, além de implicações diretas e indiretas no cenário proposto.

Os tópicos apresentados nesta seção são Governança de dados, Big Data e a correção da aplicação de ambos nas instituições públicas federais, podendo ser observados em três capítulos distintos.

### **2.1. Governança de dados**

Nesta seção é apresentado uma concepção histórica sobre Governança de dados, relatando sua real importância para o meio educacional, comercial e industrial.

Em um segundo momento, e apresentado ao leitor, os riscos de se aplicar governança de dados, principalmente quando esta aplicação não é centralizada em setores que controlem adequadamente suas diretrizes. Tais correlações são importantes para o entendimento de suas atribuições junto aos órgãos federais que será discutido posteriormente.

Para finalizar esta seção, é relatado ao leitor, as aplicações de governança de dados. Tais evidências, mostram requisitos benéficos para organizações que conseguem trabalhar suas diretrizes de forma eficiente, mostrando sua relevância ao tema proposto.

Dentre as várias situações e apresentado constantemente relações entre os tópicos e a real situação governamental dentro das diretrizes de cada tópico.

#### **2.1.1 Concepção histórica e sua relevância para o cenário**

O *Data Governance* (Governança de dados), é uma estrutura de quatro vias que inclui disponibilidade, aplicabilidade, integridade e segurança. É um conjunto de processos, usado pelas partes interessadas que usam a tecnologia, para garantir que



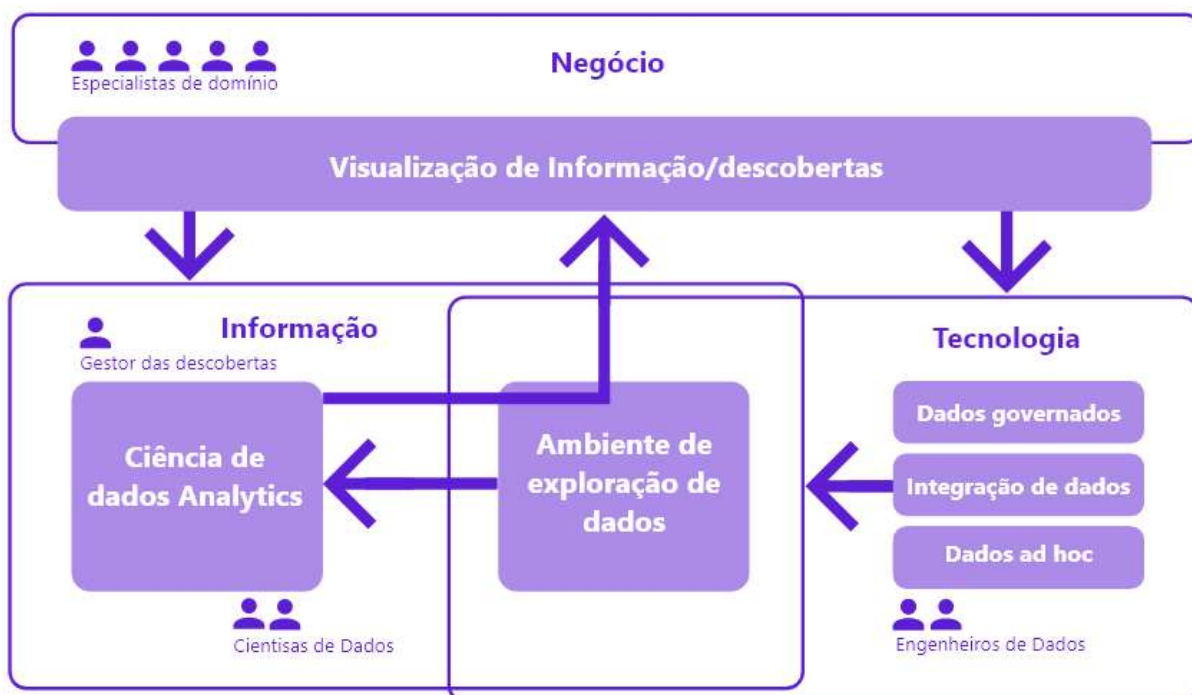
os dados importantes e críticos sejam gerenciados e protegidos (ESPINDOLA, et. al., 2017).

Envolve uma coordenação simplificada de indivíduos (pessoas), métodos (processos) e inovação (tecnologia) em uma ordem que resulta na realização do valor dos dados para qualquer organização (BARATA, 2015). Ela atua como uma ponte entre negócios e TI para decisões e iniciativas.

Embora pessoas, processos e tecnologia estejam no centro de uma boa governança de dados, a tecnologia é apenas um facilitador, não essencial para a governança de dados. Ela deve ser implementada como um fluxo de trabalho disciplinado dentro da organização. Sem essa disciplina, os dados nunca seriam tratados como uma mercadoria valiosa.

Stumpf (2017), na figura 02 abaixo, mostra a importância de Governança de dados nas organizações.

Figura 2 - Demonstrativo da utilização de governança de dados



Fonte: Stumpf (2017) – Modelo adaptado pelo autor

A Governança de dados é necessária para garantir que os ativos de informações de uma organização sejam gerenciados de maneira formal, adequada, proativa e eficiente em toda a empresa para garantir sua confiança e responsabilidade.

O *Data Governance* compreende a coleta de dados, a revisão e a padronização e a otimização do uso. Torna os dados consistentes. A Governança de Dados garante que os dados críticos estejam disponíveis no momento certo para a pessoa certa, de forma padronizada e confiável (CARVALHO, 2012). Isso infere uma melhor organização das operações de negócios. A adoção e implementação da Governança de Dados pode resultar em maior produtividade e eficiência de uma organização.

O que não se sabe, não pode ser protegido. Esse ditado é ainda mais importante para a segurança cibernética. Pesquisas líderes mostram que 65% não têm a capacidade de detectar violações de dados com o *Data Governance*, que deve ser visto como um esforço da empresa (DGI, 2017).

Pode-se implementar em segmentos, mas deve sempre e necessariamente ter uma perspectiva corporativa. Para isso, primeiro deve ser estabelecido um órgão de governança que crie a estratégia e a política necessárias para a organização. Este não é um exercício único, mas um esforço contínuo que requer monitoramento e manutenção.

O monitoramento, manutenção e revisão contínuos dos dados são cruciais e importantes. O sucesso de qualquer programa de Governança de Dados pode ser obtido através da comunicação eficaz entre TI, Unidades de Negócios e partes interessadas em dados.

Outros fatores de sucesso incluem documentação e comunicação suficientes de alterações, problemas e tarefas em um plano de comunicação simplificado e predefinido. Um programa de governança de dados seguro e de alto desempenho ajudará qualquer organização a aproveitar seus dados genuínos e dignos dos negócios com a ajuda de soluções tecnológicas apropriadas, quando necessário.

O objetivo final de qualquer programa de Governança de Dados é permitir que as organizações reajam efetivamente aos desafios e oportunidades colocados pelo mercado, adotando a filosofia “*Data First*” ( BARATA, 2015).

Como no caso de qualquer outro programa, as diferenças culturais e as abordagens isoladas para aquisição e gerenciamento de dados criam barreiras na formação e manutenção de um programa de Governança de Dados, além da falta de patrocínio comercial sênior sustentado.

Uma base inadequada e a ausência de métricas são outros motivos para uma falha na governança de dados. Uma base adequada inclui: gerenciamento de dados, modelos de dados, metadados etc. Fundamentalmente, uma base sólida do que precisa ser governado é o requisito mais básico para a governança de dados (ESPINDOLA, et. al., 2017).

Uma organização não pode ser transformada da noite para o dia, especialmente em um programa de trabalho intensivo, como o *Data Governance*. Esperar uma mudança radical de um estágio "sem programa" para um programa "em toda a empresa" em alguns meses é uma expectativa irreal e pode levar à falha no controle de dados.

Subestimar a quantidade de trabalho a ser realizado, excesso de planejamento e menos execução, falta de compromisso de linha / negócio sustentado e ausência de equipe / escritório do programa também são alguns dos motivos que contribuem para a falha na governança de dados. Planejar, definir o escopo e executar um programa de Governança de Dados pode ser desafiador, portanto pode ser feito em etapas para alcançar os resultados desejados.

### **2.1.2 Riscos aplicados a governança de dados**

Em um ambiente de negócios volátil e de mudança dinâmica, o gerenciamento de riscos e a garantia da informação desempenham um papel crítico no crescimento e sustentação gerais de qualquer organização. Devido à crescente dependência de infraestrutura aprimorada digitalmente, os CIOs enfrentam o desafio assustador de

encontrar o equilíbrio certo entre proteção e progresso, bem como privacidade e governança.

Além do dilema acima, há uma postura de segurança fragmentada e, às vezes desfocada, resultante de um mandato multitorregulatório. Aí reside a necessidade de uma estrutura holística de risco e conformidade da governança.

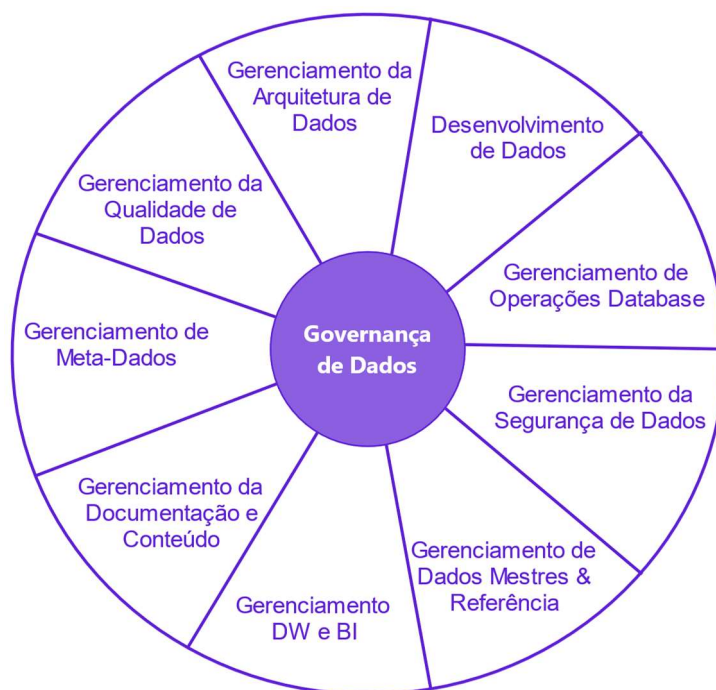
Santos (2010), destaca alguns riscos de governança e desafios de conformidade, como:

- Falta de visibilidade em dados críticos e ativos cruciais;
- Impacto e perdas devido a uma violação ou ameaça percebida;
- Cobertura incompleta de controles e ativos de segurança;
- Otimização de riscos de governança e ferramentas de gerenciamento de conformidade;
- Visibilidade do risco atual de governança e postura de conformidade;

Governança, risco e conformidade (GRC) refere-se a uma estratégia para gerenciar a governança geral de uma organização, o gerenciamento de riscos corporativos e a conformidade com os regulamentos. Pense no GRC como uma abordagem estruturada para alinhar a TI aos objetivos de negócios, enquanto gerencia efetivamente os riscos e atende aos requisitos de conformidade (SANTOS, 2010).

Stumpf (2017), mostra na figura 03 abaixo, baseando-se no modelo de DMBok, uma sugestão de uma visão sistemática para um framework de gerencia de dados.

Figura 3 – Modelo de framework para gestão de dados



Fonte: Stumpf (2017) – Modelo adaptado pelo autor

Uma estratégia de GRC bem planejada traz muitos benefícios: tomada de decisão aprimorada, investimentos em TI ideais, eliminação de silos e fragmentação reduzida entre divisões e departamentos, para citar alguns.

Gartner (2016), define que em um ambiente de TI o GRC possui três componentes principais:

- **Governança:** garantir que as atividades organizacionais, como o gerenciamento de operações de TI, estejam alinhadas de maneira a apoiar os objetivos de negócios da organização.
- **Risco:** Certifique-se de que qualquer risco (ou oportunidade) associado às atividades organizacionais seja identificado e tratado de maneira a apoiar os objetivos de negócios da organização. No contexto de TI, isso significa ter um processo abrangente de gerenciamento de riscos de TI que entra na função de gerenciamento de riscos corporativos de uma organização.

- **Conformidade:** garantir que as atividades organizacionais sejam operadas de maneira a atender às leis e regulamentos que afetam esses sistemas. No contexto de TI, isso significa garantir que os sistemas de TI e os dados contidos nesses sistemas sejam usados e protegidos adequadamente.

O cumprimento da conformidade envolve controles de TI, bem como a auditoria desses controles para garantir que eles estejam funcionando conforme o planejado. As organizações também usam controles para gerenciar riscos identificados. De fato, o termo "GRC" surgiu no início dos anos 2000, após muitos desastres financeiros corporativos altamente divulgados, o que resultou em empresas lutando para melhorar seus processos de controle interno e governança (GARTNER, 2016).

As organizações desenvolvem uma estrutura de GRC para a liderança, organização e operação das áreas de TI da organização, para garantir que elas apoiem e permitam os objetivos estratégicos da organização. A estrutura específica mensuráveis claramente definidos que iluminam a eficácia dos esforços de GRC de uma organização. Embora existam muitas boas opções de software disponíveis para ajudar a simplificar as operações de GRC, o GRC é mais do que um conjunto de ferramentas de software.

Muitas organizações consultam uma estrutura para orientação no desenvolvimento e aprimoramento de suas funções GRC, em vez de criar uma a partir do zero. Estruturas e padrões fornecem blocos de construção que as organizações podem adaptar ao seu ambiente. Pode-se destacar, o COBIT, o COSO e o ITIL são os grandes players em diversos setores.

### 2.1.3 Entendendo sobre o DAMA DMBOK: Conceitos e Diretrizes

Esta seção pode ser resumida na importância do guia DAMA-DMBOK (2018), que realmente uma grande fonte de conhecimento de diferentes áreas relacionadas ao gerenciamento de dados, reunidas e resumidas por uma forte equipe de profissionais.

O guia fornece a famosa ‘Roda DAMA’ (veja a figura 3), que explica as nove principais áreas de conhecimento. O modelo Hexágono de Fatores Ambientais é a base para a descrição de cada área de conhecimento e inclui, por exemplo, produtos, atividades, ferramentas etc. Mas ainda assim, o modelo de gerenciamento de dados DAMA tem suas próprias especificações, que consiste, no seguinte :

- Pode ser aplicada a uma ou a várias áreas do conhecimento;
- Pode ser implementada na prática, através de frameworks específicos;

Cada uma das áreas de conhecimento pertence a diferentes categorias, por exemplo, Metadados e Referência e Dados Mestres são relacionados a dados, enquanto Armazenamento de Dados e Operações têm baixo aspecto tecnológico e operacional. Então, como se deve lidar com isso? Na verdade, a ideia é cada grupo de profissionais envolvidos nos processos de governança de dados, baseando-se na Roda DAMA, encontrar a “sua própria maneira” de realizar as aplicações práticas. Ressalta-se que existem diretrizes que apoiam e norteiam tais procedimentos, porém o guia é claro com relação a sua prática.

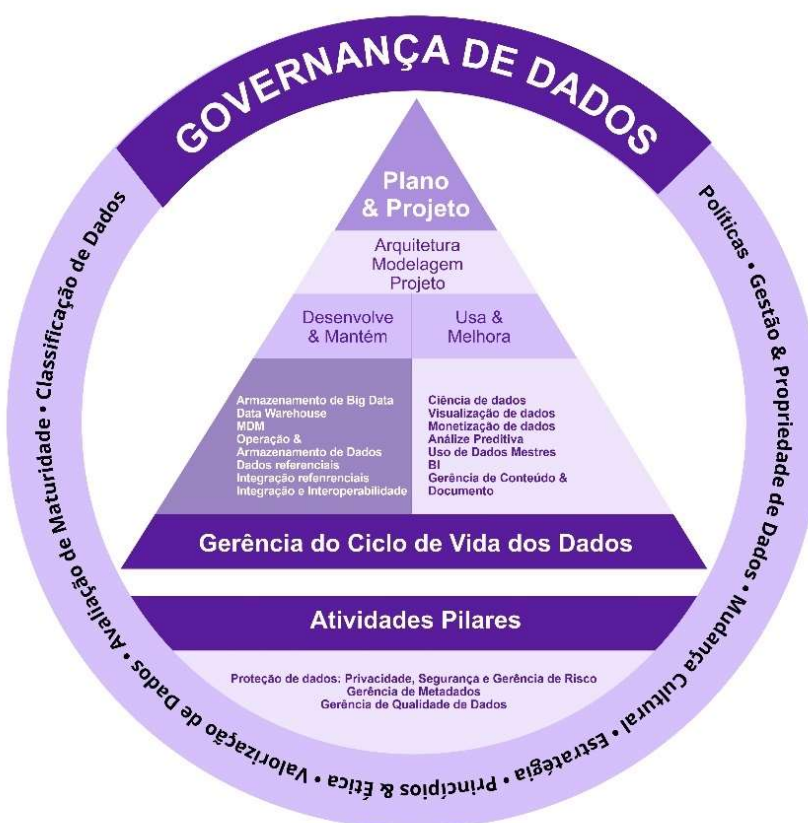
Desta forma, o guia DAMA-DMBOK (2018), ressalta que existem três pilares principais que suportam o modelo de gerenciamento de dados DAMA:

- O gerenciamento de dados é uma função de negócios, o mesmo que finanças, vendas etc.
- No escopo do DAMA, o gerenciamento de dados é considerado parte da organização de Tecnologia da Informação.

- O gerenciamento de dados é multifuncional; requer uma gama de habilidades e conhecimentos .

O assunto principal do gerenciamento de dados são os dados, O DAMA-DMBOK do ponto de vista do conteúdo classifica os dados em 4 categorias: Referência, Mestre, Transacional e Metadados. Desta forma, desperta a curiosidade de saber por que o DAMA-DMBOK considera os dados de referência e mestre como uma área de conhecimento separada e negligencia os dados transacionais.

Figura 4 – Modelo DAMA-DMBOK, segundo guia



Fonte: Diagrama DAMA DMBok V2 Evoluído – Modelo adaptado pelo autor

Metadados sendo "dados sobre dados" descrevem todos os outros tipos. Cada tipo de dado tem sua qualidade. Portanto, 'Qualidade de Dados' de acordo com



DAMA-DMBOK também é uma área de conhecimento. Arquitetura de dados e modelagem e design de dados são ferramentas que, de uma forma ou de outra, auxiliam na descrição da estrutura e do processamento de dados. A Arquitetura de Dados se concentra na documentação de fluxos de dados e cadeias de valor de dados. Data Modeling and Design é dedicado a definir requisitos de dados, fornecendo diferentes modelos de dados DAMA-DMBOK (2018).

Os dados também podem ser classificados pela forma como são armazenados. Assim, o DAMA-DMBOK apresenta uma classificação de dados estruturados e não estruturados. Os dados não estruturados ganham destaque na área de conhecimento de 'Gerenciamento de documentos e conteúdo'.

Os dados estruturados são mencionados em três áreas de conhecimento. 'Armazenamento de dados e operações' fornece informações sobre design, implementação e operação de bancos de dados. A movimentação da organização e a consolidação de dados são o núcleo de 'Integração e interoperabilidade de dados', e 'DWH e BI' se concentra em relatórios de dados.

Observando com atenção a figura 04, essas três áreas de conhecimento estão relacionadas aos aspectos tecnológicos e operacionais do processamento de dados. No modelo, todos eles estão rodeados pelo retângulo "Tecnologia". A segurança de dados é projetada pela Arquitetura e implementada por todos os blocos relacionados à tecnologia. A governança de dados é abrangente e diz respeito a processos, procedimentos e funções da função de negócios de gerenciamento de dados.

#### **2.1.4 Aplicação da governança de dados**

Por que a governança de dados é importante para as empresas? Isso pode parecer uma pergunta retórica. O poder dos dados em impulsionar as empresas hoje é bem reconhecido. Os dados são poderosos e é melhor que as ideias sejam orientadas por dados. Então, segue-se que a governança de dados também é importante. Mas por que isso é exatamente: o que a governança de dados faz para o resultado final? Uma empresa possui muitos dados recebidos externamente e mais

dados são criados ou atualizados internamente, de modo que provavelmente os dados devem ser 'governados' para que se tenha bons dados.

Ladley (2012), define que a governança de dados é um conjunto de processos que garante que ativos de dados importantes sejam formalmente gerenciados em toda a empresa. Garante que informações confiáveis sejam usadas para processos críticos de negócios, tomada de decisão e contabilidade.

Desta forma, pode-se dizer que a falta de governança de dados efetiva dentro de uma empresa garante uma coisa: a existência de dados insuficientes. Esses dados ruins se manifestam em definições inconsistentes, duplicatas, campos ausentes e outros tipos de dados clássicos.

Destaca-se os principais benefícios que por que a governança de dados é importante (LADLEY, 2012):

1. **Governança de dados economiza dinheiro:** Simplesmente, a governança de dados aumenta a eficiência. Contas duplicadas levam a esforços duplicados ou, pelo menos, levam a perda de tempo rastreando contas duplicadas em seus esforços de marketing, vendas, finanças ou analíticas. A governança de dados reduz erros no seu banco de dados, fornecendo aos negócios um banco de dados sólido para trabalhar e economiza um tempo precioso que seria usado para corrigir os dados existentes. Tempo economizado é dinheiro economizado. Além disso, a governança de dados força uma empresa a definir claramente seus dados principais e as regras que regem esses dados principais. O início de um projeto de governança de dados é uma oportunidade de ouro para colocar todos na mesma página sobre as principais definições de dados. A aplicação disso garante maior eficiência operacional ao longo do tempo.

2. **O risco da má governança dos dados:** A falta de governança eficaz dos dados é uma preocupação de segurança por duas razões: riscos externos à segurança associados a dados sujos e não estruturados e problemas de conformidade regulamentar. Dados inválidos e dados mal estruturados representam um risco à segurança pelo simples motivo de que, se tiver dados não estruturados sujos e

obstruindo o banco de dados, como poderá saber rapidamente quando algo dá errado e como monitorar com eficiência quais dados estão em risco? Boas ferramentas e práticas de governança de dados facilitam o monitoramento do que está acontecendo em seu banco de dados e facilitam a visualização de quais áreas podem estar em risco. À medida que as pessoas continuam a entender a importância de seus dados pessoais, os governos começam a tomar as maneiras pelas quais as empresas armazenam, protegem e usam os dados de seus clientes com muita seriedade.

**3. Fornecendo clareza:** Reserve um segundo para imaginar o que a garantia de dados perfeitos significaria para as organizações, e para o governo? A governança eficaz dos dados fornece a tranquilidade de que os dados geralmente são limpos, padronizados e precisos. Os efeitos disso repercutem em toda a empresa, garantindo que as métricas sejam precisas, conhecendo quais são as métricas mais importantes e principalmente como e quando aplicá-las e ainda proporcionando uma maior confiança nas análises.

Atualmente, a governança de dados é fundamental para as empresas orientadas a dados, e realmente qual empresa hoje não está tentando ser orientada a dados? Agora sabe-se por que a governança de dados é importante atualmente, e a partir desta afirmação, considere três forças principais das quais as empresas estão se alinhando para se beneficiar no futuro próximo: Internet das Coisas, Inteligência Artificial e Big Data.

Todas essas três forças trazem grandes promessas para muitas empresas por meio de uma enorme quantidade de dados, através dos quais insights e inteligência podem ser obtidos; no entanto, o influxo de dados aumenta a necessidade de um programa eficaz de governança de dados. Se uma empresa não se antecipar aos dados sujos provenientes da Internet das coisas, da IA e do Big Data, o resultado principal pode ser apenas um grande pântano de dados, em vez do aumento da inteligência e dos lucros que os membros do conselho esperavam.

Assim sendo, este objeto de estudo aborda conceitos fundamentais que auxiliam esta temática, relacionando a governança de dados aliada à uma destas três forças: o Big Data, e aplica-se através destes conceitos, diretrizes que podem e devem

ser aproveitadas nas instituições federais, de modo que garantam efetividade, objetividade, redução de custos, entre outros.

## **2.2. Big Data e suas aplicações**

Esta seção inicia-se apresentando os conceitos básicos sobre Big Data. É importante neste primeiro momento a familiarização do termo para posteriormente entender suas aplicações.

Em um segundo momento, apresenta-se as aplicações já destinadas ao cenário desta dissertação. Aqui é evidenciado não só as potencialidades de Big Data aplicadas ao setor público federal, mas também suas restrições, principalmente citando o porque elas acontecem, quais são as barreiras governamentais que impedem que Big data seja amplamente aplicado com objetividade e efetividade.

Posteriormente e citado no tópico 3 desta seção, alguns dos recursos de dados disponíveis para a Administração Pública, relatando conceitos, comparando com comissões especializadas (como por exemplo a comissão europeia).

No quarto tópico são apresentados requisitos técnicos e não técnicos relacionados a aplicabilidade no setor público federal. As diretrizes apontadas são registros do Senado Federal, constituído em 2012 que evidenciam e contribuem para esta temática.

No quinto e ultimo tópico desta seção, é apresentado a importância de Big Data para a Administração Pública Federal. Questões relacionadas à sua contribuição direta e indireta, bem como os benefícios que trás ao governo são mensurados nesta seção.

### **2.2.1 Conceito histórico e atribuições**

Big Data é um termo que descreve o grande volume de dados em velocidade de geração e transmissão elevadas e diferentes formatos - estruturados e não estruturados - que inunda uma empresa no dia-a-dia. Mas não é a quantidade de dados que é importante. É o que as organizações fazem com os dados que

interessam. O Big Data pode ser analisado para obter informações que levam a melhores decisões e movimentos estratégicos de negócios (SILVA, 2015).

O uso do Big Data está se tornando comum pelas empresas para superar seus pares. Na maioria das indústrias, os concorrentes existentes e os novos entrantes usarão as estratégias resultantes dos dados analisados para competir, inovar e capturar valor.

O Big Data ajuda as organizações a criar novas oportunidades de crescimento e categorias novas de empresas que podem combinar e analisar dados do setor. Essas empresas possuem amplas informações sobre produtos e serviços, compradores e fornecedores, preferências do consumidor que podem ser capturadas e analisadas.

Embora o termo “Big Data” seja relativamente novo, o ato de coletar e armazenar grandes quantidades de informações para análise final tem muito tempo. O conceito ganhou força no início dos anos 2000, quando o analista do setor Doug Laney articulou a definição atual de Big Data (Figura 5) como os cinco Vs:

Figura 5 - Demonstração de Big Data através dos 5 Vs.

Volume	Variedade	Velocidade	Veracidade	Valor
 <p>Quantidade de dados que são gerados a cada segundo</p>	 <p>Tipos de dados que podem ser utilizados</p>	 <p>Velocidade a qual os dados são gerados e transmitidos</p>	 <p>Grau de organização e confiabilidade dos dados</p>	 <p>Necessidade de se transformar dados em valor</p>
Terabytes Registros Transações Arquivos	Estruturado Semiestruturado Não estruturado	Tempo real Streams Batch	Fidedignidade Autenticidade Origem Relevância Disponibilidade	Estatística Eventos Correlações Hipotética

Fonte: Tehseen (2016), modelo adaptado pelo autor.

De acordo com Dias (2018) e Oracle (2019), a Figura 1 acima pode ser definida da seguinte maneira:

**Volume:** as organizações coletam dados de várias fontes, incluindo transações comerciais, mídias sociais e informações de dados de sensores ou de máquina para máquina. No passado, armazená-lo teria sido um problema - mas as novas tecnologias (como o *Hadoop*) diminuíram o fardo. O nome 'Big Data' em si está relacionado a um tamanho enorme. O tamanho dos dados desempenha um papel muito crucial na determinação do valor dos dados. Além disso, se um dado específico pode ser considerado um Big Data ou não, depende do volume de dados. Portanto, '**Volume**' é uma característica que precisa ser considerada ao lidar com 'Big Data'.

**Variedade.** Os dados são fornecidos em todos os tipos de formatos - desde conjuntos de dados estruturados, dados numéricos em bancos de dados tradicionais a documentos de texto não estruturados, e-mail, vídeo, áudio, dados de cotações de ações e transações financeiras. Variedade refere-se a fontes heterogêneas e à natureza dos dados, estruturados e não estruturados.

**Velocidade:** os dados são transmitidos a uma velocidade sem precedentes e devem ser tratados em tempo hábil. Tags RFID, sensores e medição inteligente estão aumentando a necessidade de lidar com *Torrents* de dados em tempo quase real.

O termo '**velocidade**' refere-se à velocidade de geração de dados. A rapidez com que os dados são gerados e processados para atender às demandas determina o potencial real nos dados.

O Big Data *Velocity* lida com a velocidade com que os dados fluem de fontes como processos de negócios, *logs* de aplicativos, redes e sites de mídia social, sensores, dispositivos móveis etc. O fluxo de dados é massivo e contínuo.

**Veracidade:** Um dos pontos mais importantes de qualquer informação é que ela seja verdadeira. Com o Big Data não é possível controlar por exemplo cada *hashtag* do Twitter ou notícia falsa na internet, mas com análises e estatísticas de grandes volumes de dados é possível compensar as informações incorretas.

**Valor:** O último V é o que torna Big Data relevante: tudo bem ter acesso a uma quantidade massiva de informação a cada segundo, mas isso não adianta nada se não puder gerar valor. É importante que empresas entrem no negócio do Big Data, mas é sempre importante lembrar dos custos e benefícios e tentar agregar valor ao que se está fazendo.

Nos dias anteriores, planilhas e bancos de dados eram as únicas fontes de dados consideradas pela maioria dos aplicativos. Os dados na forma de e-mails, fotos, vídeos, dispositivos de monitoramento, PDFs, áudio etc. também estão sendo considerados aplicativos de análise. Essa variedade de dados não estruturados coloca certos problemas para armazenamento, mineração e análise de dados (ORACLE, 2019).

Desta forma, a Oracle (2019) destaca os benefícios do Big Data, como:

- as empresas podem utilizar inteligência externa enquanto tomam decisões;
- melhor atendimento ao cliente;

- identificação precoce de risco para o produto/serviços, se houver;
- melhor eficiência operacional.

A importância do Big Data não gira em torno da quantidade de dados que uma empresa possui, mas de como ela utiliza os dados coletados. Toda empresa usa dados à sua maneira; quanto mais eficiente a empresa usa seus dados, maior o potencial para crescer. A empresa pode pegar dados de qualquer fonte e analisá-los para encontrar respostas que poderão permitir os itens de 1 a 7 descritos a seguir (ORACLE, 2019).

1. **Economia de custos:** algumas ferramentas de Big Data, como *Hadoop* e *Cloud-Based Analytics*, podem trazer vantagens de custo para os negócios quando grandes quantidades de dados devem ser armazenadas e essas ferramentas também ajudam na identificação de maneiras mais eficientes de fazer negócios.

2. **Reduções de tempo:** a alta velocidade de ferramentas como o *Hadoop* e a análise na memória podem identificar novas fontes de dados, o que ajuda as empresas a analisar dados e a tomar decisões rápidas com base nos aprendizados.

3. **Entender as condições do mercado:** analisando Big Data, pode-se entender melhor as condições atuais do mercado. Por exemplo, analisando o comportamento de compra dos clientes, uma empresa pode descobrir os produtos mais vendidos e produzir produtos de acordo com essa tendência. Com isso, ele pode ficar à frente de seus concorrentes.

4. **Controlar a reputação online:** as ferramentas de Big Data podem fazer análises de sentimentos. Portanto, pode-se obter *feedback* sobre quem está dizendo o que diz respeito à sua empresa. Se desejar monitorar e melhorar a presença *on-line* dos negócios, as ferramentas de Big Data podem ajudar nisso.

5. **Usar o Big Data Analytics para aumentar a aquisição e retenção de clientes:** o cliente é o ativo mais importante de que qualquer empresa depende. Não existe um negócio único que possa reivindicar sucesso sem primeiro ter que estabelecer uma base sólida de clientes. No entanto, mesmo com uma base de



clientes, uma empresa não pode se dar ao luxo de desconsiderar a alta concorrência que enfrenta. Se uma empresa demora a aprender o que os clientes estão procurando, é muito fácil começar a oferecer produtos de baixa qualidade. No final, resultará em perda de clientela, e isso cria um efeito geral adverso no sucesso dos negócios. O uso de Big Data permite que as empresas observem vários padrões e tendências relacionados ao cliente. Observar o comportamento do cliente é importante para gerar lealdade.

**6. Usar o Big Data Analytics para resolver o problema dos anunciantes e oferecer informações de marketing:** a análise de Big Data pode ajudar a mudar todas as operações comerciais. Isso inclui a capacidade de corresponder às expectativas dos clientes, mudar a linha de produtos da empresa e garantir que as campanhas de marketing sejam poderosas.

**7. Análise de Big Data como motor de inovações e desenvolvimento de produtos:** outra vantagem do Big Data é a capacidade de ajudar as empresas a inovar e reconstruir seus produtos.

Os melhores exemplos de Big Data podem ser encontrados tanto no setor público quanto no privado. Desde publicidade direcionada, educação e indústrias já mencionadas (assistência médica, seguros, manufatura ou serviços bancários), há cenários da vida real, em atendimento a clientes ou entretenimento. Até o ano 2021, 1,7 megabytes de dados serão gerados a cada segundo para cada pessoa no planeta; o potencial de crescimento organizacional orientado a dados no setor de hospitalidade é enorme (FAGUNDES et al., 2017).

Enfatizando a aplicação de Big Data no setor bancário, pode-se imaginar, que quantidade de dados nos setores estão subindo de forma rápida a cada segundo. Desta forma, a aplicação de Big Data, pode ajudar a detectar:

- o uso indevido de cartões de crédito;
- uso indevido de cartões de débito;
- tratamento de risco de crédito de risco;

- clareza de negócios;
- alteração nas estatísticas do cliente;
- lavagem de dinheiro;
- mitigação de riscos.

### 2.2.2 Aplicações potenciais de Big Data para a Administração Pública Federal e suas restrições

Quatro aplicações potenciais para o setor público foram descritas e desenvolvidas em Zillner et al. (2013) por demonstrar o uso de tecnologias de Big Data no setor público.

Tabela 1 - Critérios dos cenários de aplicação para a Administração Pública.

Resumo dos cenários de aplicação para a Administração Pública	
Nome	Monitoramento e supervisão de atividades regulamentadas para operadores de jogos de azar on-line
Sumário	Grandes volumes de dados disponíveis dificultam a regulação e a supervisão eficientes das atividades
Sinopse	Monitorar os operadores de jogos de azar on-line para o controle das atividades regulamentadas e a detecção de fraudes. O usuário deste aplicativo é o órgão público responsável pela atividade de supervisão. Este procedimento é uma obrigação regulatória da administração pública; os operadores de jogos de azar on-line devem fornecer as informações ao público regulador por meio de um canal de comunicação específico. Os dados em tempo real são recebidos dos operadores de jogos a cada 5 minutos.
Objetivos de Negócios	Garantir a conformidade com os regulamentos, prevenção e detecção de fraudes e investigação criminal.
Nome	Eficiência operacional em agência de trabalho
Sumário	Extrair valor de grandes volumes disponíveis de dados não utilizados
Sinopse	Ative uma nova gama de serviços personalizados, melhore o atendimento ao cliente e reduza os custos de operação na agência federal do trabalho alemã. Todos os trabalhadores desempregados estavam recebendo os mesmos serviços padrão, apesar de terem perfis diferentes. Os dados históricos de seus clientes foram analisados, incluindo perfis, intervenções e o tempo necessário para encontrar um

	emprego. Com base nessa análise, a segmentação de clientes foi desenvolvida.
Objetivos de Negócios	Reduza o custo e melhore a qualidade do serviço: agora eles conseguem encontrar um novo emprego em um período mais curto.
Nome	Segurança pública em cidades inteligentes
Sumário	Grandes volumes de dados disponíveis a partir de sensores, mídias sociais e chamadas de emergência podem ser combinados para fornecer segurança pública eficaz.
Sinopse	Cidades inteligentes equipadas com sensores e infraestruturas de comunicação ajudam o setor público a manter as cidades e seus cidadãos seguros. Ter informações precisas e atualizadas permite respostas melhores e mais rápidas durante emergências e resulta em menos danos e vítimas. As fontes típicas para obter essas informações podem vir de chamadas de resposta a emergências, câmeras de vigilância e forças móveis (como um carro-patrolha da polícia) que chegaram ao local. Nos últimos anos, as mídias sociais mostraram um potencial interessante para a coleta de informações que auxiliam na obtenção de uma imagem precisa da consciência situacional (van Kasteren et al. 2014). Todas as informações coletadas são coletadas em um centro de comando e controle onde um operador pode decidir como orientar as forças móveis disponíveis.
Objetivos de Negócios	Resposta rápida a emergências, prevenção de danos e menos vítimas.
Nome	Policimento preditivo usando dados aberto
Sumário	Reutilização de dados públicos abertos para fornecer policiamento preditivo
Sinopse	Governos de todo o mundo iniciaram iniciativas de dados abertos para disponibilizar dados do setor público em prol da transparência e permitir que terceiros ofereçam serviços com base nesses dados. Um desses serviços pode ser descrito como policiamento preditivo, onde dados históricos de crimes são usados para descobrir automaticamente tendências e padrões. Os padrões identificados ajudam a obter insights sobre os problemas relacionados ao crime que uma cidade está enfrentando e permitem um destacamento mais eficaz e eficiente das forças policiais (Wang et al. 2013; PredPol 2013).
Objetivos de negócios	Diminuição significativa do crime, uso eficiente de forças móveis.

---

Wang et al. (2013) menciona que, para a Big Data na Administração Pública, as restrições podem ser resumidas como descrito a seguir.

**Falta de vontade política:** de fazer o setor público tirar proveito dessas tecnologias. Isso exigirá uma mudança de mentalidade dos altos funcionários do setor público.

**Falta de pessoas qualificadas e orientadas para os negócios:** cientes de onde e como a Big Data pode ajudar a solucionar os desafios do setor público e quem pode ajudar a preparar a estrutura regulatória para o desenvolvimento bem-sucedido de soluções de Big Data.

**O novo Regulamento Geral de Proteção de Dados e as diretivas PSI exibem algumas incertezas:** sobre o impacto na implementação de iniciativas de Big Data e dados abertos no setor público. Assim, os dados abertos devem ser um catalisador do setor público para o setor privado para estabelecer uma poderosa indústria de dados.

**Ganhando impulso na adoção:** hoje, há mais marketing sobre Big Data no setor público do que experiências reais para aprender quais aplicativos são mais rentáveis e como devem ser implantados. Isso requer o desenvolvimento de um conjunto padrão de soluções de Big Data para o setor.

**Numerosos órgãos da administração pública:** (naqueles amplamente descentralizados), muita energia é perdida e permanecerá assim até que uma estratégia comum seja realizada para a reutilização de plataformas entre tecnologias.

### 2.2.3 Recursos de dados disponíveis para a Administração Pública

De acordo com Zillner et al. (2013), sobre a reutilização de informações do setor público, a informação do setor público (PSI) é definida da seguinte forma: abrange qualquer representação de atos, fatos ou informações – e qualquer compilação de tais atos, fatos ou informações – qualquer que seja seu meio (escrito em papel ou armazenado em formato eletrônico ou como gravação sonora, visual ou audiovisual), realizado por órgãos públicos.

Um documento mantido por um órgão do setor público é um documento em que o órgão do setor público tem o direito de autorizar a reutilização. Segundo Correia (2004), quanto à disponibilidade das informações produzidas por esses órgãos públicos, e na ausência de diretrizes específicas, o organismo produtor é livre para decidir como disponibilizá-las: de forma direta aos usuários finais, estabelecendo um público / parceria privada ou terceirizar a exploração comercial dessas informações a operadores privados.

Desta forma, Correia (2004) esclarece que as atividades que não se enquadram na tarefa pública: “normalmente incluem o fornecimento de documentos produzidos e cobrados em caráter comercial e em concorrência com outros no mercado”.

Sobre a natureza do PSI disponível, existem várias abordagens. O Livro Verde sobre o PSI (Comissão Europeia 1998) propõe algumas classificações, como:

- distinção PSI entre administrativa e não administrativa;
- distinção do PSI em relação à sua relevância para o público;

Além disso, ele pode ser diferenciado de acordo com seu valor potencial de mercado e, em alguns casos, de acordo com o conteúdo dos dados pessoais:

- distinção PSI de acordo com seu anonimato;

A quantidade mais importante de dados produzidos pelo setor público é textual ou numérica, em comparação com outros setores, como a assistência médica, que produz uma quantidade de imagens eletrônicas. Como resultado das iniciativas de governo eletrônico dos últimos 15 anos, parte desses dados é criada em formato digital, 90% de acordo com a McKinsey (MCKINSEY GLOBAL INSTITUTE, 2011).

De acordo com uma pesquisa realizada para a formulação da Parceria Europeia de Valorização de Big Data para representantes do setor público (ZILLNER et al., 2013), o principal ativo de dados apontados pelos autores é todo o sistema de

setor público, registros, Bancos de Dados e Sistemas de Informação, dos quais os mais significativos são:

- cidadãos, empresas e propriedades (por exemplo, registros de base, transações);
- dados fiscais;
- dados de segurança;
- gerenciamento de documentos, e as crescentes transações eletrônicas;
- compras e despesas públicas;
- órgãos e funcionários públicos;
- dados geográficos relacionados a cadastros;
- conteúdo relacionado à cultura, educação e turismo;
- documentos legislativos;
- dados estatísticos (dados socioeconômicos que poderiam ser usados pelo setor privado);
- dados geoespaciais.

#### **2.2.4 Requisitos da administração pública: Técnicos e não técnicos.**

A agregação de dados por meio dos limites administrativos de maneira não baseada em solicitação é um desafio real, pois essas informações podem revelar informações pessoais e de segurança altamente confidenciais quando combinadas com várias outras fontes de dados, não apenas comprometendo a privacidade individual, mas também a segurança civil.

Os direitos de acesso aos conjuntos de dados necessários para uma operação devem ser justificados e obtidos. Quando uma nova operação é realizada com dados existentes, é necessário obter uma notificação ou uma licença da Agência de Privacidade de Dados.

O anonimato deve ser preservado nesses casos, para que a dissociação de dados seja necessária. Preocupações individuais com privacidade e segurança pública devem ser tratadas antes que os governos possam ser convencidos a

compartilhar dados mais abertamente, não apenas publicamente, mas compartilhando de maneira restrita com outros governos ou entidades internacionais.

Outra dimensão é a regulamentação para o uso da computação em nuvem, de forma que o setor público possa confiar nos provedores de nuvem. Além disso, a falta de fornecedores europeus de computação em nuvem de Big Data no mercado europeu também é uma barreira para adoção.

Faltam cientistas e tecnólogos especializados em dados que possam capturar e processar essas novas fontes de dados. Quando as tecnologias de Big Data se tornam cada vez mais adotadas nos negócios, os profissionais qualificados de Big Data se tornam mais difíceis de encontrar.

Os órgãos públicos podem se distanciar com as habilidades que já possuem, mas precisam garantir que essas habilidades avancem. Além das pessoas com orientação técnica, há uma falta de conhecimento nas pessoas com orientação comercial que estão cientes do que a Big Data pode fazer para ajudá-las a resolver os desafios do setor público.

Desta forma, o Senado Federal (2012) explica que os requerimentos não técnicos incluem:

- disposição de fornecer e adotar tecnologias de Big Data e também de saber usá-lo;
- necessidade de abordagens (políticas) nacionais ou europeias comuns - como as políticas nacionais de interoperabilidade e dados abertos. Falta de liderança neste campo;
- uma incompatibilidade geral entre inteligência de negócios em geral e Big Data em particular no setor público.

Já em relação aos requisitos técnicos o Senado Federal (2012) destaca como descrito nos parágrafos a seguir.

**Descoberta de padrões:** identificação de padrões e semelhanças para detectar comportamentos criminais ou ilegais específicos no cenário de aplicação de monitoramento e supervisão de operadores de jogos de azar on-line (e também para cenários de monitoramento semelhantes no setor público). Esse requisito também é aplicável no cenário para melhorar a eficiência operacional na agência de trabalho e no cenário de policiamento preditivo.

**Compartilhamento de dados/integração de dados:** necessário para superar a falta de padronização dos esquemas de dados e a fragmentação da propriedade dos dados. Integração de múltiplas e diversas fontes de dados em uma plataforma de Big Data.

**Informações em tempo real:** ative a análise de dados novos/em tempo real para tomada de decisão instantânea, para obter informações em tempo real a partir dos dados.

**Segurança e Privacidade de Dados:** procedimentos legais e meios técnicos que permitem o compartilhamento seguro de dados e a preservação da privacidade. As soluções para esse requisito podem desbloquear o amplo uso de Big Data no setor público. Os avanços na proteção e privacidade dos dados são fundamentais para o setor público, pois podem permitir a análise de grandes quantidades de dados pertencentes ao setor público sem divulgar informações confidenciais. Esses problemas de privacidade e segurança estão impedindo o uso de infraestruturas de nuvem (processamento, armazenamento) por muitos órgãos públicos que lidam com dados confidenciais.

**Transmissão de dados em tempo real:** como a capacidade de colocar sensores está aumentando em cenários de aplicativos de cidades inteligentes, há uma alta demanda por transmissão de dados em tempo real. Será necessário fornecer recursos distribuídos de processamento e limpeza para sensores de imagem, a fim de não reduzir os canais de comunicação e fornecer apenas as informações necessárias para a análise em tempo real, que fornecerá sistemas de conscientização situacional para os tomadores de decisão.



**Análise de linguagem natural:** extraia informações de fontes on-line não estruturadas (por exemplo, mídias sociais) para permitir a mineração de sentimentos. Reconhecimento de dados de entradas de linguagem natural, como texto, áudio e vídeo.

**Análise preditiva:** conforme descrito no cenário do aplicativo para policiamento preditivo, onde o objetivo é distribuir forças e recursos de segurança de acordo com a previsão de incidentes, forneça previsões com base no aprendizado de situações anteriores para prever a alocação ideal de recursos para serviços públicos.

**Modelagem e Simulação:** ferramentas específicas do domínio para modelagem e simulação de eventos de acordo com dados de eventos passados para antecipar os resultados de decisões tomadas para influenciar as condições atuais em tempo real, por exemplo, em cenários de segurança pública.

Prossegue-se agora ao Capítulo 3, no qual se estuda por meio da pesquisa de campo o cenário objeto da pesquisa com vistas a responder os questionamentos principais estudados nesta dissertação.

### **2.2.5 A importância do Big Data na Administração Pública Federal**

Por trás da política da montanha-russa que balança países como Estados Unidos, Alemanha, Japão, Coreia e também o Brasil, uma revolução muito mais silenciosa, mas igualmente importante, está em andamento. Os gerentes governamentais estão avançando no uso de “big data” e estão tendo um impacto. É o centro de um esforço importante para transformar a saúde do governo federal e melhorar os resultados dos programas federais, como argumentou um recente relatório da Academia Nacional de Administração Pública (NAOPA, 2018).

Desta forma, indaga-se sobre qual o problema relacionado ao Big Data? É tentador olhar para a revolução silenciosa dos dados como apenas o próximo passo pequeno e lógico na medição do desempenho do governo, uma continuação da evolução em curso que começou com a Revisão de Desempenho Nacional da Al Gore.

Depois disso, veio uma série de agendas de administração, cada uma delas focada (de maneiras muito diferentes) na produção de melhores informações para obter melhores resultados. E seria fácil dizer que a coisa da big data é apenas uma melhoria incremental.

Desta forma (NAOPA, 2018), retrata que o papel do Big Data na administração pública federal possui um papel importante, onde o fornecimento de dados explodiu, com muito mais dados de muito mais fontes. O governo não tem escolha sobre adotar a revolução da big data. Tudo o que o governo faz, desde o custo dos cuidados de saúde até o estado do meio ambiente, é de forma simples inundado de números, como nunca antes.

O movimento é grande demais para alguém ignorar. Não apenas o governo está produzindo seus próprios dados em um volume cada vez maior, como também estão surgindo dados do setor privado no setor privado, tanto através do investimento de empresas privadas quanto do crescimento explosivo das mídias sociais. Mesmo que o governo quisesse ignorar a revolução dos dados, não poderia - os dados estão chegando de todos os lugares.

Mais dados estão disponíveis em tempo real. O famoso programa CompStat da cidade de Nova York começou porque os comandantes da polícia perceberam que não precisavam esperar até o final do ano para obter os números dos crimes. Eles começaram a coletar estatísticas de crimes. Mas agora, através de tudo, desde câmeras de rua a informações pessoais, funcionários do governo em todos os lugares podem ver o que está acontecendo, aqui e agora.

Mais dados são granulares, conectados às operações reais dos programas. Os Centros de Controle e Prevenção de Doenças sabem como as vacinas contra a gripe estão afetando o surto de gripe, e a Agência de Proteção Ambiental pode ver o impacto das regras do ar limpo na qualidade do ar - incluindo as emissões de instalações individuais. Os aplicativos para smartphone rastreiam todos os voos que voam para qualquer lugar, com a possibilidade de planejar em tempo real. É possível detalhar da política ampla aos resultados reais e acompanhar o que está acontecendo.

Pela primeira vez, funcionários no topo - no Gabinete de Administração e Orçamento e nos escritórios dos secretários de gabinete - podem ver os mesmos dados em tempo real, ao mesmo tempo, dos gerentes de linha de frente. Esta é uma oportunidade emergente não totalmente desenvolvida. Mas a capacidade dos altos funcionários de entender o que está acontecendo em escritórios individuais e de pedir que os gerentes desses escritórios respondam inevitavelmente transformará a administração federal.

Mais desses dados estão vivendo separadamente dos mundos anteriores da análise de políticas (como análise de custo-benefício e avaliação de programas) e gerenciamento de desempenho (como avaliações em larga escala dos resultados do programa). Esse tsunami de dados está vivendo uma vida separada - e crescente -. Ele se baseia em perguntas diferentes, usando dados diferentes, processados por meio de ferramentas diferentes, fornecendo respostas diferentes.

Esses dados e análises de big data estão possibilitando fazer perguntas antes inimagináveis. Quais regulamentos do governo estão nos livros há mais tempo? Quais regulamentos estão mais conectados com os outros? Um estudo recente de big data da Deloitte descobriu que 12% de todas as seções do Código de Regulamentos Federais não foram tocadas desde pelo menos a década de 1970 - e que 67% não foram editadas desde que foram publicadas. Existem 17.800 seções que são correspondências extremamente próximas com outras seções, diferindo apenas em poucas palavras. Se estivermos interessados em simplificar as regulamentações federais, a análise de big data fornece uma maneira de saber para onde procurar. Isso simplesmente não era possível antes (DELOITTE, 2019).

A revolução dos dados não apenas fornece muito mais informações sobre o que está acontecendo. Também cria novas maneiras de comunicar o que está acontecendo. Há muito debate sobre os “desertos alimentares” nas cidades - áreas onde as pessoas têm pouco acesso a alimentos saudáveis e não têm um bom transporte para alcançar melhores lugares para comprá-los.

Um estudo de Baltimore produziu um mapa fascinante de desertos alimentares lá. Ele coloca quatro bancos de dados diferentes um sobre o outro para produzir uma análise clara e convincente de um problema muito complicado. Uma coisa é produzir equações de regressão múltipla sofisticadas que exploram as interações de variáveis importantes, mas de uma forma que ninguém, a não ser o conhecimento, pode digerir. Outra coisa é desenhar um mapa que capte questões importantes em bairros individuais, de uma forma que faça sentido para todos. Ver é acreditar, e acreditar torna mais fácil agir sobre isso (BALTIMORE, 2019).

A própria natureza da revolução da big data pode ajudar a superar o problema fundamental de tantos esforços anteriores para levar uma melhor análise ao governo: tanto investimento em tantos estudos que poucas pessoas leem e que produziram tão pouco resultado. Um bom mapa - sobre desertos alimentares, por exemplo - pode identificar problemas e ajudar a impulsionar soluções. A transparência das informações dificulta a ignição. Se pudermos descobrir quais regulamentos federais são quase idênticos e se sobrepõem, há um eleitorado natural para consertar isso.

Essa revolução facilita o ataque ao problema ainda mais fundamental de oferta e demanda: produzir respostas para as perguntas dos formuladores de políticas, em vez de tentar convencê-los a fazer as perguntas sobre as quais temos dados. Com tantas informações, tão potencialmente digeríveis e importantes, pode ser muito mais fácil dar aos formuladores de políticas o que eles querem, quando querem, em um idioma que eles possam entender - e agir.

Não é um trabalho fácil. De fato, uma das maiores limitações é que simplesmente não existem muitos funcionários do governo especializados em mineração de dados e análise de dados. Outra é que parte desse trabalho exige que as pessoas que sabem como detalhar, perguntem o que está acontecendo, por que isso importa e o que fazer a respeito. De fato, a revolução dos dados não é tanto sobre números, mas também sobre capital humano: levar as pessoas que sabem como impulsionar a revolução.

Mas não é como se o governo tivesse uma escolha. A sociedade, em geral, está galopando para a frente, e os cidadãos não terão muita paciência com um governo que não acompanha. A boa notícia, porém, é que a revolução dos dados está proporcionando ao governo enormes novas oportunidades. O mundo está constantemente procurando a próxima coisa. No governo, é isso. A revolução dos dados não apenas tem o potencial de melhorar a saúde das organizações governamentais. Também pode melhorar a eficácia do governo.

## **2.3. Governança de dados aliado ao Big Data**

Nesta seção, aborda-se sobre a junção entre governança de dados aliado ao Big Data. O porque são cruciais para uma organização e como podem ser aproveitados dentro das Instituições Federais, são apenas alguns dos pontos explorados aqui.

### **2.3.1 Porque governança de dados e crucial para ambientes de Big Data**

Organizações em todo o mundo estão investindo em sistemas capazes de hospedar e processar dados de maneiras nunca antes imaginadas. Em alguns casos, as empresas estão até reformulando seus ambientes de TI existentes com base nesses novos sistemas. Esses sistemas de big data produziram resultados tangíveis: aumento de receita e custos mais baixos. No entanto, os resultados positivos estão longe de ser garantidos. Para realmente obter valor dos dados, essas novas plataformas devem ser governadas (HENRIQUES & COSTA, 2018).

O termo governança de dados causa medo nos corações de muitos profissionais de dados. Como muitas vezes é vagamente definido e mal compreendido, muitos simplesmente recorrem a uma abordagem apenas de tecnologia para resolver suas necessidades de governança. A complexidade que vem com muitos sistemas de big data torna essa abordagem baseada em tecnologia especialmente atraente, embora seja sabido que apenas a tecnologia raramente será suficiente. O que talvez seja menos conhecido é que as próprias tecnologias devem ser revisadas ao otimizar a governança de dados hoje (ROSSOUW, 2012).

A governança de dados não é a linhagem, administração ou gerenciamento de dados mestre. Cada um desses termos é frequentemente ouvido em conjunto com - e mesmo no lugar de - governança de dados. Na verdade, essas práticas são componentes dos programas de governança de dados de algumas organizações. São componentes importantes, mas são meramente componentes (STUMPF, 2017).

No essencial, a governança de dados trata de gerenciar formalmente dados importantes em toda a empresa e, assim, garantir que o valor seja derivado. Embora os níveis de maturidade variem de acordo com a organização, a governança de dados geralmente é alcançada por meio de uma combinação de pessoas e processos, com a tecnologia usada para simplificar e automatizar aspectos do processo.

Cita-se, por exemplo, segurança. Mesmo os níveis básicos de governança exigem que os ativos de dados importantes e sensíveis de uma empresa sejam protegidos. Os processos devem impedir o acesso não autorizado a dados confidenciais e expor todos ou parte desses dados a usuários com uma "necessidade de saber" legítima.

As pessoas devem ajudar a identificar quem deve ou não ter acesso a determinados tipos de dados. Tecnologias como sistemas de gerenciamento de identidades e recursos de gerenciamento de permissões simplificam e automatizam os principais aspectos dessas tarefas.

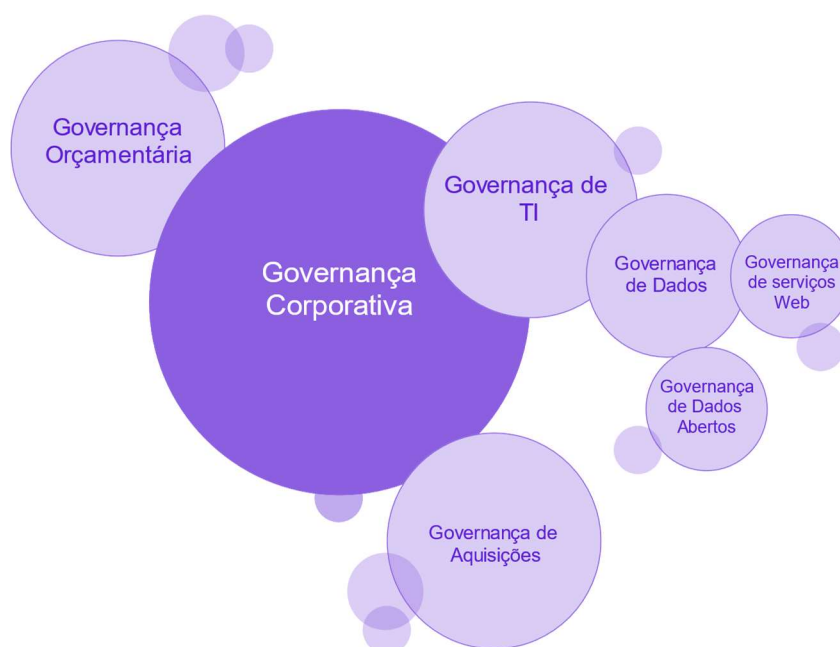
Algumas plataformas de dados simplificam ainda mais as tarefas, vinculando os registros existentes baseados em nome de usuário / senha, como o Active Directory, e permitindo maior expressividade ao atribuir permissões, além dos relativamente poucos graus de liberdade proporcionados pelos bits do modo POSIX.

Também deve-se reconhecer que, à medida que a velocidade e o volume dos dados aumentarem, será quase impossível para humanos (por exemplo, administradores de dados ou analistas de segurança) classificá-los em tempo hábil. Às vezes, as organizações são forçadas a manter novos dados bloqueados em uma célula de retenção até que alguém os classifique e exponha adequadamente aos usuários finais.

Tempo valioso é perdido. Felizmente, os provedores de tecnologia estão desenvolvendo maneiras inovadoras de classificar dados automaticamente, diretamente quando ingeridos ou logo depois. Ao alavancar essas tecnologias, um pré-requisito essencial do processo de autorização é atendido, minimizando o tempo para o *insight* (STUMPF, 2017).

Stumpf (2017), enfatiza através da figura 06 a integração entre as governanças (de Tecnologia da Informação, Orçamentária, Dados) e deixa um leque para aplicações de Big Data.

Figura 6 – Governança de dados integrado com o ambiente corporativo



Fonte: Stumpf (2017) – modelo adaptado pelo autor

Assim sendo, argumenta-se: Qual a diferença entre a governança de dados na era do big data? Fazendo referencia ass 5V's do Big Data, governar esses sistemas pode ser complicado. As organizações geralmente são forçadas a agrupar *clusters* separados, cada um com seu próprio objetivo comercial ou armazena e processa tipos de dados exclusivos, como arquivos, tabelas ou fluxos.

Mesmo se a costura em si for feita com cuidado, as lacunas serão rapidamente expostas, pois a proteção consistente dos conjuntos de dados em vários repositórios pode ser extremamente suscetível a erros.

Arquiteturas convergentes simplificam bastante a governança. Em sistemas convergentes, vários tipos de dados (por exemplo, arquivos, tabelas e fluxos) são integrados em um único repositório de dados que pode ser controlado e protegido de uma só vez. Não há costura a ser feita em si, porque todo o sistema é cortado e governado contra o mesmo tecido.

Além dos Vs, há outra diferença mais sutil. A maioria das distribuições de big data, se não todas, inclui uma amálgama de diferentes mecanismos de análise e aprendizado de máquina situados "em cima" do (s) repositório (s) de dados. *Spark* e *Hive* são apenas dois dos mais populares em uso hoje. Essa flexibilidade é ótima para os usuários finais, porque eles podem simplesmente escolher a ferramenta mais adequada às suas necessidades específicas de análise.

O problema da perspectiva de governança é que essas ferramentas nem sempre respeitam os mesmos mecanismos ou protocolos de segurança, nem registram ações de forma completa, consistente ou em repositórios que podem ser dimensionados - pelo menos não "prontos para uso" (ROSSOUW, 2012).

Como resultado, os profissionais de big data podem ser flagrados ao tentar atender às demandas de conformidade ou de auditoria sobre, por exemplo, linhagem de dados - um componente de governança que visa responder perguntas como: "De onde esses dados vieram e o que aconteceu?" com o tempo? "

### **2.3.2 O papel integral da Governança de dados em ambientes de big data**

A governança de dados é um conceito multifacetado, mas fornece as ferramentas e os processos para promover o entendimento dos dados em toda a empresa. É um programa abrangente, não um projeto, e deve incluir um conjunto principal de soluções para fornecer uma base de governança adequada (YIU, 2012).



Essas soluções incluem um glossário de negócios, dicionários de dados e linhagem de dados para definir dados, termos e atributos de negócios, bem como fontes de dados, uso, relacionamentos e interdependências. A governança de dados também deve atribuir claramente a responsabilidade e a propriedade entre os envolvidos, administradores e proprietários de dados, bem como um mecanismo para gerenciar consultas e resolver problemas.

Historicamente, a governança de dados tem sido intimamente associada à garantia da conformidade regulatória e, embora isso seja verdade, o papel da governança de dados é muito mais amplo na era do big data. Por exemplo, o gerenciamento de metadados é uma parte crucial da governança e os metadados desempenham um papel importante para as organizações descobrirem *insights* analíticos (TEHSEEN, 2016).

A governança de dados também desempenha um papel crítico nos esforços de qualidade de dados, à medida que as organizações continuam lutando para avaliar, melhorar e relatar a qualidade de seus dados.

Os desafios que a governança de dados abrangente de hoje ajuda as organizações a superar - os de acessibilidade, usabilidade, significado e qualidade - todos aumentam exponencialmente no mundo do big data. Os ambientes de big data são potenciais tesouros para *insights*, mas sem governança, responsabilidade e colaboração e suporte organizacional adequados, eles podem ser buracos negros de dados não utilizados (SILVA, 2015).

A chave para governar esses ambientes é gerenciar e definir esses dados em toda a cadeia de suprimento de dados - um esforço que começa quando os dados são ingeridos nas organizações e entra em qualquer ambiente interno, seja um *data warehouse* ou um *data lake*, e continua ao longo dos dados ciclo da vida. Em toda a cadeia de fornecimento de dados, questões importantes precisam ser abordadas. Esses incluem:

Transparência e rastreabilidade são elementos-chave que podem ser rastreados por meio de metadados e linhagem de dados - de onde vieram os dados, por quais processos e sistemas foram movidos na organização e como eles foram transformados? A qualidade dos dados é uma preocupação constante, pois os dados são considerados para análise (SILVA, 2015).

Com possíveis transformações, o que sabe-se sobre esses dados e em que pode-se confiar? É preciso, consistente e confiável? Os usuários de negócios podem depender disso para gerar análises e insights precisos? Acessibilidade e entendimento são fundamentais para usuários de negócios. Sem eles, é como ter um armazém cheio de equipamentos e materiais, mas nenhuma chave para a porta e nenhuma instrução sobre como usar o que está dentro - nada será construído ou o que será construído provavelmente desmoronará.

Desta forma, pode-se indagar: Quais dados estão disponíveis? Como é usado internamente? Como é definido e quais são os termos comerciais relacionados? Essas definições diferem entre as linhas de negócios ou departamentos? Os dados devem ser claramente categorizados, organizados e disponíveis para os usuários e bem definidos, para que os dados certos sejam selecionados para a tarefa correta.

Propriedade e colaboração são críticas. Não basta saber de onde vieram os dados e o que são; é preciso haver responsabilidade e prestação de contas contínuas por esses ativos. Quem possui os dados? Os proprietários e administradores de dados devem ser claramente definidos para que os usuários de negócios tenham recursos a quem recorrer, para perguntas sobre uso e aplicabilidade (SILVA, 2015).

Um programa abrangente de governança de dados responderá a todas essas perguntas e fornecerá uma estrutura sólida para que os dados organizacionais extraídos sejam confiáveis, compreensíveis e utilizáveis. Não fazer isso pode resultar em decisões de negócios baseadas em dados incompletos ou incorretos, que podem custar caro tanto em receita quanto em perda de reputação, produtividade e oportunidades perdidas.

A governança de dados está ganhando importância não apenas devido ao crescente volume e velocidade de dados e ao surgimento de ambientes de big data, mas também devido às crescentes complexidades regulatórias e ao desafio incansável de garantir a qualidade dos dados para produzir resultados de qualidade. E embora a governança tenha sido historicamente relegada à TI e à conformidade, são os líderes de negócios de hoje que precisam aproveitar dados e análises para obter vantagem competitiva e melhorar os resultados (SILVA, 2019).

Enfrenta-se uma democratização dos dados no sentido de que os dados não são mais apenas uma preocupação da TI, pois as restrições de recursos e as demandas de negócios direcionam uma crescente necessidade de empoderamento dos usuários de negócios por meio de recursos de autoatendimento. As empresas precisam ter dados na ponta dos dedos para resolver problemas de negócios, para que possam transformar rapidamente esses dados em insights acionáveis.

É por isso que a governança de dados é tão crítica na era do big data. A necessidade e o uso de dados estão cada vez mais nas mãos dos usuários de negócios, o que significa que precisa haver uma estrutura sólida que defina todos os aspectos dos dados e sua utilização (SILVA, 2019).

### **2.3.3 Como as diferenças de Big Data podem afetar a Governança de dados**

Embora o Big Data e a governança de dados já existam há algum tempo, o relacionamento entre eles ainda é uma novidade. A necessidade de governança de dados em big data surge da necessidade das empresas de obter informações precisas, confiáveis e acionáveis sobre os dados existentes. Na ausência de uma política adequada de governança de dados, os projetos de dados podem resultar em custos excessivos e em informações enganosas, o que acabaria por causar danos irrevogáveis aos negócios.

De acordo com (AUTOR, ANO), a governança de big data deve começar definindo uma política geral de governança de dados. Isso incluiria políticas claras

definidas para inventário de dados, propriedade dos dados, qualidade dos dados, segurança das informações e retenção de dados.

A existência de políticas para cada uma delas permite que a empresa identifique as partes interessadas certas e aplique procedimentos para garantir a eficiência e a confiabilidade, enquanto reduz os riscos de falha. Em seguida, é necessário definir os padrões para os dados, começando pelos conjuntos de dados críticos que derivam mais valor para os negócios.

Isso inclui identificar elementos de dados críticos, definir metadados e relacionamentos e, finalmente, documentá-los para referência futura. Deve haver um processo de automação para lidar com grandes cargas de dados provenientes de diferentes fontes da organização, pois os processos manuais não podem atender a uma variedade tão grande de informações (AUTOR, ANO).

Hoje, os modelos analíticos são críticos para uma plataforma de big data, portanto, andam de mãos dadas com uma política de governança de dados. O modelo analítico deve ser simples o suficiente para ser entendido e utilizado por usuários de negócios e analistas de dados. Além disso, deve haver um processo para atender aos novos requisitos analíticos dos usuários finais. Como pode ser observado, o advento de ferramentas mais novas como o *SlamData*, os modelos analíticos se tornam menos críticos, no entanto, a necessidade de governança não diminui (AUTOR, ANO).

Finalmente, e mais importante, deve haver uma equipe dedicada para lidar com as responsabilidades da governança de big data. A equipe deve consistir de um patrocinador, diretor de dados e vários especialistas no assunto para monitorar, revisar, revisar e aprimorar constantemente a qualidade do processo e as informações.

Em resumo, não se espera que o Big Data dê seu verdadeiro valor a uma empresa sem uma política de governança de dados estabelecida. A governança de dados é uma novidade no que diz respeito ao big data, no entanto, está cada vez mais chamando a atenção das equipes de big data. Em pouco tempo, ele se tornará

obrigatório para qualquer projeto de big data, permitindo que as empresas aproveitem totalmente seu poder.

## **2.4. Aplicações de governança de dados e Big Data em Instituições públicas federais: Uma visão globalizada**

Uma boa governança de dados pode contribuir para definir uma visão comum; melhorar a implementação e coordenação coerentes; e fortalecer as bases institucionais, regulatórias, de capacidade e técnicas para controlar e gerenciar melhor o ciclo de valor dos dados, ou seja, coletar, gerar, armazenar, proteger, processar, compartilhar e reutilizar dados, como forma de aumentar a confiança e agregar valor (SILVA, 2019).

A boa governança de dados é imprescindível para governos que pretendem se tornar mais orientados a dados como parte de sua estratégia digital. Pode ajudar a extrair valor dos ativos de dados, permitindo maior acesso, compartilhamento e integração de dados no nível organizacional e além, além de aumentar a eficiência e a responsabilidade geral. No entanto, embora o conceito não seja novo, a maioria dos governos estão lutando para colocá-lo em prática.

Barc (2019), menciona que a governança de dados é cada vez mais relevante para as práticas de proteção de dados em escala global, de maneira mais exclusiva e explícita. No entanto, uma abordagem forte e desequilibrada da superproteção de dados pode reduzir o valor do compartilhamento de dados, como na prestação de serviços públicos transfronteiriços.

Recentemente, o uso indevido de dados por empresas privadas e as crescentes preocupações dos cidadãos sobre o gerenciamento de dados no setor público provocaram uma intervenção do governo para melhorar a proteção de dados pessoais (OCDE, 2019). Como resultado, o uso ético e transparente dos dados está agora no topo da agenda política.

Os fluxos de dados aumentaram nas organizações, setores (por exemplo, empresa para governo) e fronteiras, adicionando outro nível de complexidade à

governança de dados em um mundo globalizado e interconectado. A governança de dados não é mais uma questão limitada aos limites organizacionais, mas uma preocupação multinacional resultante do compartilhamento de dados além-fronteiras.

Nesse contexto, instrumentos internacionais, como o Regulamento Geral de Proteção de Dados, procuraram “devolver aos cidadãos o controle sobre seus próprios dados” (OCDE, 2019) e tomar medidas internacionais para impedir o uso indevido de dados. O regulamento geral sobre proteção de dados avançou a agenda de proteção de dados, subjacente à necessidade de estruturas comuns para garantir a proteção de dados além-fronteiras.

No entanto, a superproteção de dados pode resultar do mal-entendido dos regulamentos nacionais e internacionais e conduzir a mudanças em termos de abordagens políticas (por exemplo, de abertura por padrão para "abrir se possível, protegido se necessário"). O desafio global neste estágio é garantir o equilíbrio certo entre fluxos de dados gratuitos e proteção de dados (OCDE, 2019).

Embora os países da OCDE tenham frequentemente definido elementos relevantes para a governança de dados do setor público no contexto do governo digital, dados abertos, gerenciamento de dados e / ou estratégias e / ou políticas de IA, esses elementos são frequentemente fragmentados. Em alguns cenários, essa desconexão está profundamente enraizada nos intrincados arranjos de governança que apoiam essas políticas (por exemplo, diferentes organizações do setor público que lideram essas políticas ou falta de clareza em termos de liderança e responsabilidades), colocando, portanto, barreiras importantes para a integração e compartilhamento de dados (BARC, 2019).

Uma governança holística de dados também pode ajudar a habilitar o governo como plataforma (uma das principais dimensões de um governo digital). Por exemplo, o desenvolvimento de ferramentas de dados comuns, mas flexíveis (por exemplo, plataformas de compartilhamento de dados) fornece soluções que podem ser reutilizadas em todo o setor público (ALGMIN & ZAID, 2018).

Em um nível mais técnico, a fragmentação também resulta de desafios herdados em termos de qual organização gera e controla os dados e a impossibilidade de compartilhar e acessar esses dados à luz de acordos legais específicos, levando a políticas e soluções técnicas isoladas que aumentam a impossibilidade de construir um governo integrado e conectado.

A falta de um modelo abrangente de governança de dados pode levar à proliferação ou duplicação de padrões de dados e soluções técnicas para compartilhamento de dados, dificultando a interoperabilidade de dados entre diferentes organizações e setores e afetando a possibilidade de integração de dados, processos e organizações. Também poderia levar a várias solicitações para os cidadãos fornecerem os mesmos dados pessoais várias vezes ao setor público desnecessariamente (ALGMIN & ZAID, 2018).

Uma estrutura de governança de dados deve garantir o gerenciamento adequado dos dados durante todo o seu ciclo de vida (Ghavami, 2015). Por exemplo, nos últimos anos, o movimento aberto de dados do governo permitiu uma discussão mais aprofundada da necessidade de fortalecer a liderança e administração de dados no setor público.

Isso também abriu uma discussão mais técnica sobre práticas aprimoradas de gerenciamento de dados, por exemplo, em torno da produção, armazenamento, processamento e compartilhamento para uma maior abertura de dados. No entanto, esses elementos não foram entendidos como parte de esforços mais amplos de dados do setor público que conectam todos os estágios do ciclo de valor dos dados.

Os países repentinamente perceberam o valor da catalogação de dados para fins de abertura e descoberta, mas falharam em reconhecer como essas iniciativas também traziam benefícios políticos relevantes para a produtividade no setor público. Por outro lado, em alguns países da OCDE, uma cultura bem estabelecida de eficiência do setor público levou ao desenvolvimento de registros de dados como um meio de melhorar o compartilhamento interinstitucional de dados.

No entanto, essa mentalidade ofuscou o crescente valor da abertura de dados do governo e do envolvimento e colaboração com atores externos para encontrar soluções para os desafios das políticas. Como resultado, os países que outrora lideraram o antigo movimento de governo eletrônico (com forte foco na eficiência) ficaram muito atrás daqueles que duplicaram os esforços para compartilhar e abrir dados aos usuários como forma de promover negócios e inovação social (GHAVAMI, 2015).

Países da OCDE, como Canadá, Irlanda, Holanda, Reino Unido e Estados Unidos, avançaram ou estão caminhando para a definição de estratégias abrangentes de dados como meio de aumentar a coesão do setor público e promover a integração de políticas e ferramentas (OCDE, 2019).

Essas estratégias abrangem a maioria, se não todos, os estágios do ciclo de valor dos dados do governo (da produção de dados e sua proteção à abertura e reutilização dos dados). Ainda assim, cada estágio requer arranjos específicos, pois eles produzem benefícios políticos específicos (por exemplo, dados abertos permitem o uso de dados como uma plataforma para maior envolvimento e colaboração do usuário, e melhores práticas de produção de coleta de dados podem ajudar a reduzir o viés político).

A OCDE observou que um forte foco em questões técnicas, como principal resultado da governança de dados, pode desencaminhar decisões políticas relacionadas a dados. Por exemplo, concentrando-se principalmente na adoção de soluções tecnológicas, como interfaces de programação de aplicativos (APIs) e padrões de dados, em vez de também permitir a organização organizacional, governança e cultural adequadas contextos para tornar essas ferramentas valiosas para enfrentar os desafios políticos. Todos esses são elementos-chave da boa governança de dados (OCDE, 2019).

Em alguns casos, os países da OCDE investiram recursos para definir papéis estratégicos (por exemplo, administradores de dados, diretores de dados) para apoiar a governança de dados através da definição de um tecido institucional mais forte. O



estabelecimento dessas funções estratégicas pode ajudar no dimensionamento e na sustentação da implementação de políticas e na construção de uma maior maturidade de dados no setor público (OCDE, 2018).

Isso ocorreu no contexto de estratégias de dados ou políticas de dados abertos [por exemplo, Coréia e Estados Unidos. No entanto, na maioria dos países, a liderança e / ou administração de dados ainda são incompreendidos, limitando a governança de dados às atividades do departamento de TI e não como um fator que pode ajudar a alcançar as metas da política por meio de melhores práticas de gerenciamento e compartilhamento de dados.

### **3. METODOLOGIA E APLICAÇÃO DA PESQUISA**

#### **3.1. Compreendendo as etapas metodológicas**

Esta dissertação foi desenvolvida em várias etapas metodológicas. A primeira pode-se dizer que é o embasamento do capítulo destinado a revisão da literatura, onde foi explorado através de pesquisas diretas (livros e obras autorais) e indiretas (teses de doutorado, mestrado, artigos, revistas, jornais, etc.) para compor o entendimento temático em questão.

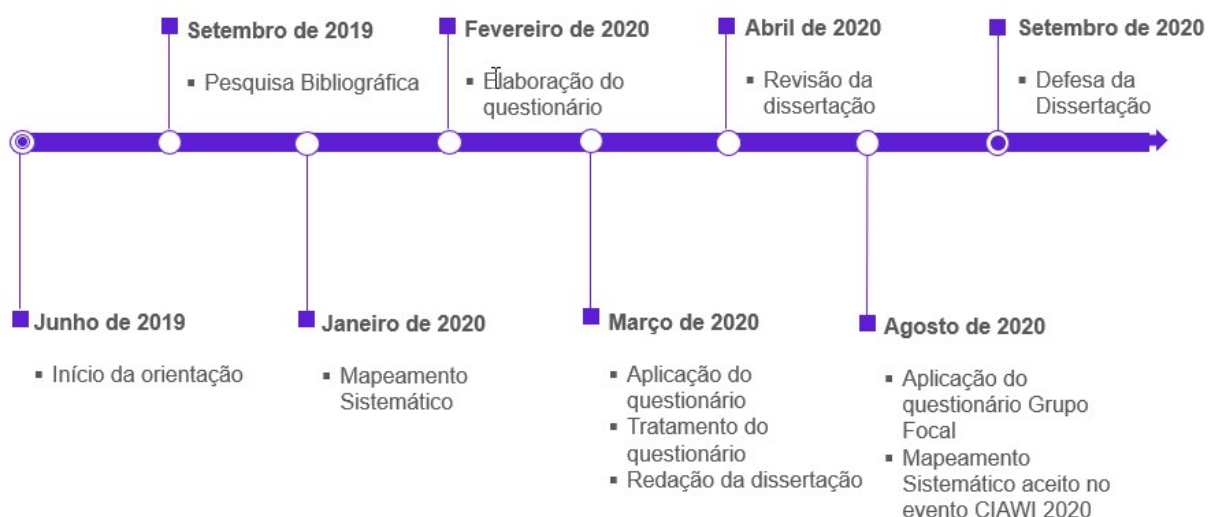
Em um segundo momento, apresenta-se o mapeamento sistemático (APÊNDICE - A), cujo o estudo antecede esta dissertação, que também é de punho acadêmico. O artigo foi aprovado para publicação na 17ª Conferência Ibero-Americana, Lisboa – Portugal, em dezembro de 2020.

Na próxima etapa, foi desenvolvido uma pesquisa do tipo survey (APÊNDICE - B). Esta pesquisa foi fundamental para a compreensão e entendimento das aplicações de Big Data nas Instituições Públicas Federais. Tal pesquisa foi disponibilizada pelo Google Forms durante um período e seu link enviado aos principais órgãos brasileiros. O relato dos participantes foi registrado e pode ser conferido na seção “Resultados e Discussão”, destinado a pesquisa correlata.

Abaixo, na figura 7, é possível verificar todas as etapas desta dissertação, onde fica compreendido o início das atividades relacionadas ao período de orientação

científica, pesquisa bibliográfica, mapeamento sistemático, pesquisa survey, pesquisa do grupo focal, confecção da guideline e conclusão dos estudos.

Figura 7 – Compreensão das etapas metodológicas.



Fonte: Elaborado pelo autor.

Posteriormente, foi realizada uma pesquisa de grupo focal (APÊNDICE - C). Tal pesquisa foi destinada a um grupo de especialistas em governança de dados e big data. Os procedimentos foram embasados na literatura de (DEVENPORT, 1954), que serviu de guia e fundamentação para a formulação do conjunto de perguntas em anexo.

Por fim, baseando-se na pesquisa de grupo focal, foi proposto uma Guideline como parte colaborativa deste objeto de estudo. Tal contribuição, trás relatos entre diretrizes e regras para que futuras instituições possam utilizar-se, para implantação das boas práticas de Governança de dados e Big Data.

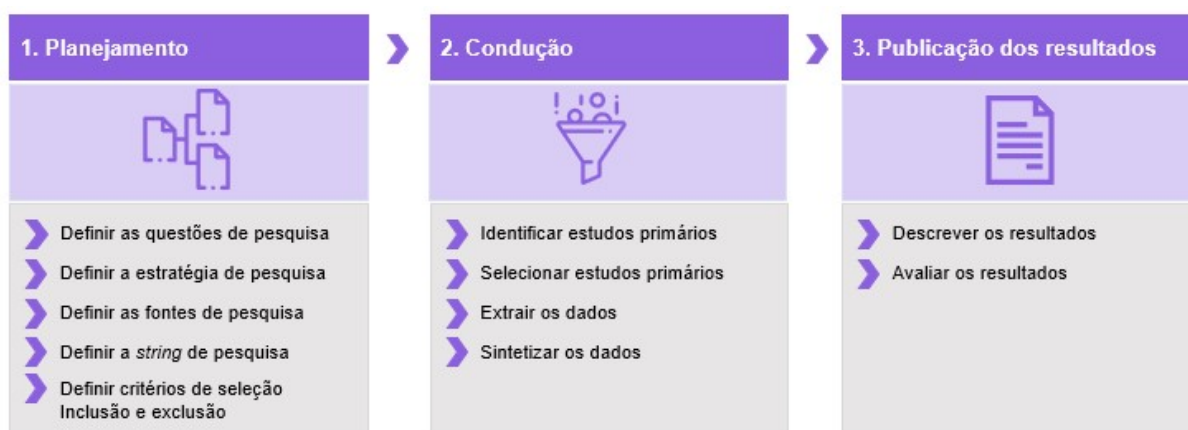
### 3.2. Apresentação do Mapeamento Sistemático

O artigo (APÊNDICE - A) apresenta o mapeamento sistemático da literatura referente uma Análise do Impacto das Práticas de Big Data na Administração Pública Federal - APF, a qual busca obter subsídios para responder quais os objetivos e maiores desafios quando da implantação de Big Data e/ou Ecossistemas de Big Data.

O artigo é antecessor à esta dissertação, e contribuiu diretamente para o engajamento e aprofundamento desta pesquisa. O mapeamento sistemático está fortemente relacionado com o conhecimento da literatura correlata que pode auxiliar no entendimento da questão em epígrafe.

Na figura 08 abaixo, ajuda na compreensão das etapas que constituiu o mapeamento sistemático que antecedeu à esta dissertação.

Figura 8 – Etapas do mapeamento sistemático.



Fonte: Elaborado pelo autor.

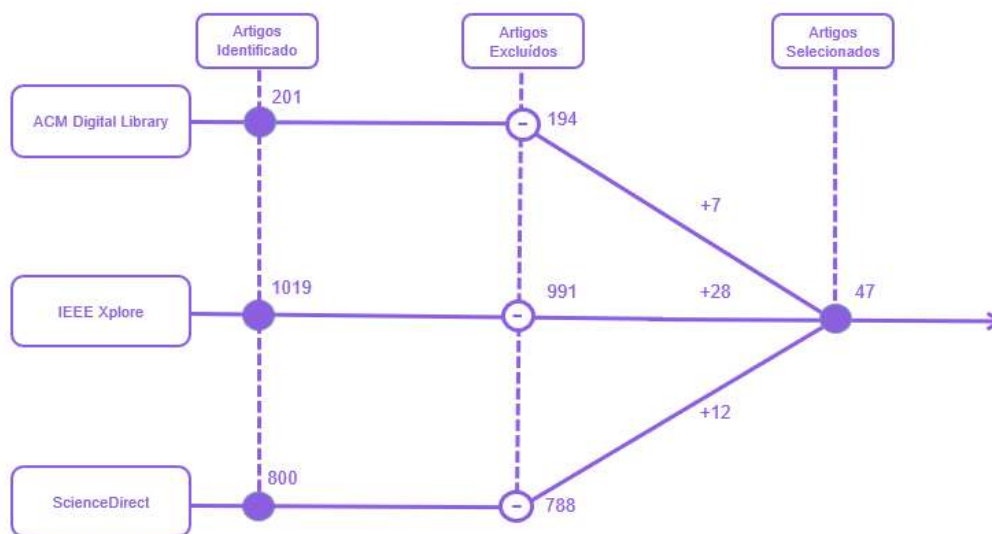
As pesquisas iniciais realizadas no artigo, contribuem diretamente para a compreensão das diretrizes de implantação de Big Data no setor público, que precisa estar perfeitamente alinhado com a missão institucional do órgão que se propõe a implantar tais diretrizes, mas que os gestores públicos nem sempre percebem suas instituições como aptas a implantarem projetos de Big Data.

Desta forma surgiu a necessidade de aplicar o questionário referente a pesquisa survey, pois entendia-se que por mais que se tenha investimentos públicos

para a tecnologia em questão, a prática estava longe de ser ideal em várias instituições que participaram do processo.

Na figura 09, fica registrado a síntese dos resultados aplicados junto aos procedimentos referentes a pesquisa de cunho bibliográfico.

Figura 9 – Síntese dos procedimentos de pesquisa bibliográfica.



Fonte: Elaborado pelo autor.

### 3.3. Apresentação da Pesquisa Survey

Para este estudo, foi escolhida a metodologia de pesquisa bibliográfica para embasamento de pesquisa do tipo *survey*.

A pesquisa bibliográfica foi fundamental para o entendimento não só dos conceitos fundamentais atinentes ao Big Data, como também das motivações do uso do Big Data e de tecnologias relacionadas em órgãos públicos. Como evidenciado ao longo de todo o presente texto, a maioria dos artigos encontrados, em especial aqueles que se relacionam com a utilização de Big Data por organizações governamentais estão ligados a casos fora do Brasil.

A discussão detalhada da bibliográfica foi objeto da seção “Referencial teórico” e se encontram sintetizadas, ainda, no mapeamento sistemático que antecedeu a presente dissertação, da qual cópia pode ser visualizada no Apêndice A.

Neste sentido, compreendeu-se a necessidade de uma pesquisa de campo. A pesquisa de tipo *survey* consiste no método de coleta de informações diretamente de pessoas a respeito de seus conhecimentos e vivência profissional no foco da pesquisa (BOCCATO et al., 2006).

A partir dos objetivos traçados e do público alvo que deseja atingir elaboram-se as perguntas. A partir disso deve ser elaborada uma estratégia de ação que deve permear a elaboração da maioria das perguntas. Estas indagações estão ilustradas na Figura 10.

Figura 10 – A Concepção e a Função do Questionário



Fonte: Moscarola (1990), modelo adaptado pelo autor.

Para a realização de tal pesquisa, e com respostas das quais se pudessem tomar alguma validade de conclusões ou de direcionamentos de conclusões, seria necessário que o respondente ou entrevistado:

- tivesse alguma relação de trabalho com a área de TI;
- tivesse conhecimento mínimo sobre Big Data e termos relacionados;
- trabalhasse prestando serviços dentro de órgãos públicos, ainda que sendo trabalhadores contratados em regime de terceirização;
- tivesse algum conhecimento acerca do órgão público para o qual prestassem serviço, bem como dos resultados ou projetos relacionados com o uso do Big Data.

Logo, qualquer questionário que fosse utilizado na pesquisa deveria coletar estas informações juntamente com as demais, para que se aferisse algum grau de pertinência do conjunto de resposta sobre o Big Data no serviço público coletadas.

Algumas das perguntas aludiram a questões de volumetria de dados guardados, uma vez que parte do público que trabalha com Big Data pode entender como Big Data apenas operações sobre conjuntos de dados à partir de determinado volume. Esta dissertação, contudo, não procurou deixar isso claro para os entrevistados, uma vez que essa volumetria que comporia o conceito do Big Data encontra variação na literatura.

As demais perguntas procuram aludir às questões principais que esta dissertação procura auxiliar a responder. Cumpre relembrar ao leitor que o objetivo principal da pesquisa é que pela análise de práticas e aplicações de Big Data no setor público brasileiro, em âmbito federal, se entenda quais os objetivos e maiores desafios e dificuldades relacionados com a adoção da Big Data.

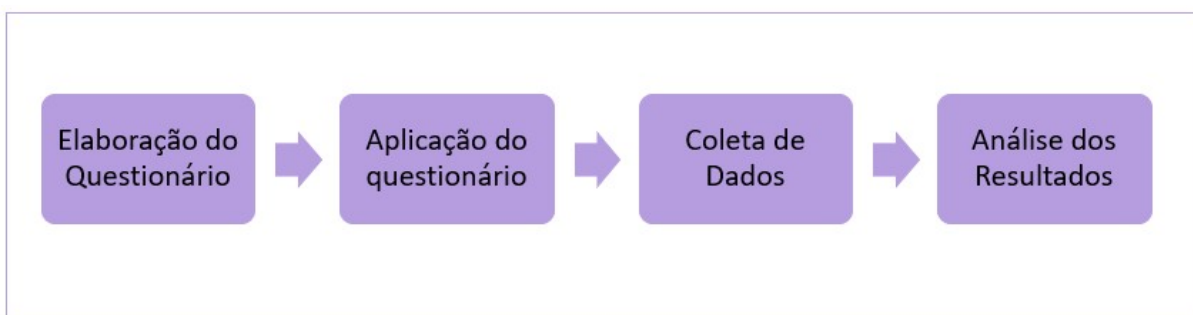
É necessário, ainda, destacar os objetivos específicos, os quais consideram um escopo mais focado em aspectos relacionados com as necessidades e com os papéis dos órgãos públicos mediante a sociedade administrada, o que se busca inferir com base nas questões elencadas a seguir.

- Quais são os impactos econômicos das soluções de Big Data no setor público?
- Quais as oportunidades e desafios da adoção de Big Data no setor público?
- Quais os impactos das práticas de Big Data no setor público?

Com base em questões que envolvessem estas quatro perguntas ligadas aos objetivos geral e específicos da dissertação seria possível aferir o alinhamento do caso da Administração Pública Brasileira com os casos aferidos em outros países, bem como a adequação dos projetos de Big Data nos órgãos públicos às suas respectivas missões institucionais, algo que é considerado condição fundamental de casos de sucesso, conforme visualizado na literatura correlata.

Em adição, seria possível visualizar, como está sendo motivada a implantação e utilização do Big Data e que resultados podem já ser compreendidos, sempre com foco nos resultados principais que se espera de organizações ligadas ao governo, em especial questões de eficiência e eficácia. A Figura 11 demonstra as etapas básicas do método utilizado.

Figura 11 – Etapas da Pesquisa.



Fonte: Elaborado pelo autor.

A seguir se apresenta as etapas acerca da pesquisa de campo:

1. **Elaboração do Questionário:** além das premissas apresentadas em epígrafe, as questões foram elaboradas com base no conhecimento e experiência profissional de casos observados no dia a dia de uma área de tecnologia relacionada a aplicações para Big Data, focando no setor público federal. O autor deste trabalho exercia a função de construtor do formulário e sua divulgação por meio da plataforma Google Forms, atendendo todos os requisitos de segurança e normalização institucional. As perguntas do questionário foram elaboradas abordando questões como: o cargo no setor público, qual o poder de atuação (Executivo, Legislativo ou Judiciário), além de questões relevantes sobre TI, experiências, conhecimento sobre Big Data, diretrizes, políticas de segurança para o armazenamento da informação, entre outras.
2. **Aplicação do Questionário:** a aplicação do questionário foi feita por meio de link de acesso para o público alvo escolhido para realização da pesquisa. O formulário foi enviado para profissionais ligados à tecnologia Big Data (sempre

focando em Setores Públicos Federais) de diferentes regiões e portes, proporcionando uma maior abrangência e alcance na coleta dos dados.

3. **Coleta de Dados:** foi utilizada a ferramenta on-line de formulário do Google Forms com perguntas descritivas com opções de respostas para a escolha (Apêndice A).
4. **Análise dos Resultados:** a elaboração do questionário faz-se necessário para a compreensão dos processos e diretrizes que norteiam a gestão de políticas de armazenamento e segurança para Big Data, buscando por meio das perguntas o foco na utilização, políticas, sistemas, segurança, armazenamento, entre outros.

### 3.2.1 Etapas da aplicação do questionário

Buscando uma compreensão na literatura, May (2004) relata que o questionário é somente um conjunto de questões realizado para gerar dados com o intuito de atingir objetivos específicos de um projeto. Chagas (2000) menciona que a aplicação de questionários é de suma importância para a obtenção de conhecimentos específicos em áreas diferentes e confeccionar um questionário não é uma tarefa das mais fáceis, já que além do esforço, envolve tempo e bastante conhecimento sobre a proposta.

Diante de tais concepções este objeto de estudo limitou-se na formulação de um questionário com 33 questões, conforme Apêndice A, cumprindo as seguintes etapas:

**1ª Etapa:** disponibilização do Termo de Consentimento Livre e Esclarecido (TCLE), que busca a conscientização do participante de maneira livre e espontânea, informando sobre os objetivos do questionário, levantamento de dados e informações relevantes sobre os processos;

**2ª Etapa:** um campo obrigatório que disponibiliza ou não o formulário para ser preenchido pelo participante, uma pergunta direta se deseja ou não participar do mesmo.



**3ª Etapa:** esta terceira etapa, explora-se as informações básicas do participante como cargo que ocupa na empresa (Administração Pública), em qual sistema de poder trabalha, qual órgão, etc.

**4ª Etapa:** compreende-se em um conjunto de informações básicas relativas ao emprego do Big Data em si. Informações de como familiarização com Big Data, experiência na área, se os órgãos públicos estão estruturados para o Big Data, entre outros.

**5ª Etapa:** compreende-se em um conjunto de questões mais abrangentes relacionadas a Big Data, enfatizando políticas, normas, regimentos, sistemas de segurança e possíveis tecnologias que poderão ser acopladas ao sistema.

Conforme explanado anteriormente, o meio utilizado para a realização desta pesquisa foi através do Google Forms. O fato da plataforma ser uma das mais utilizadas atualmente, os serviços da Google são acessados e conhecidos pela área acadêmica, técnica e científica, permitindo assim uma melhor disseminação dos formulários, estatísticas geradas por meio dos questionamentos (perguntas diretas de múltipla escolha), e possibilitando o compartilhamento dos resultados com todos aqueles que participaram diretamente da pesquisa.

A Google Forms passou a ser utilizada por acadêmicos, cientistas, etc., por possuir uma interface intuitiva, simples e de fácil compreensão. Para quem realiza o questionamento (formatação das perguntas), a plataforma oferece várias opções de controle, além de permitir a geração de gráficos que facilitam a compreensão dos resultados.

A pesquisa foi disponibilizada durante 14 dias, contendo um número de 83 participantes, porém 11 registros foram excluídos por inconsistência nos dados, distribuídos em 45 órgãos públicos, dos quais todos aceitaram a participar da mesma. Deste modo, avaliou-se a participação dos envolvidos mediante os questionamentos obtendo dados precisos sobre as diretrizes de gestão de tecnologias, políticas e diretrizes de segurança voltadas para Big Data.

### 3.4. Apresentação da pesquisa Grupo Focal

Através da pesquisa survey, realizada na seção anterior, foi possível perceber a necessidade de se realizar uma nova pesquisa, mas focada e com especialistas que estão vivenciando na prática os conceitos desta dissertação.

Para modificar comportamentos é necessário primeiro que se entenda o contexto no qual ocorrem os significados e a importância a ele atribuído pelos seus agentes. Enquanto pequena parcela do significado do contexto comportamental é de cunho pessoal e particular, grande parte deste é culturalmente moldado e socialmente construído, (GILBERT, 1991).

Mediante esta necessidade, surgiu então o interesse pela pesquisa de grupo focal.

Desta forma, apesar das análises preliminares realizadas nas instituições públicas federais, foi percebido a necessidade de se buscar algo mais específico para o entendimento das ações e diretrizes para aplicação de governança de dados aliando a big data, e ficou compreendido assim, a necessidade da pesquisa de grupo focal.

A pesquisa foi realizada com um grupo específico de profissionais, contendo 7 integrantes, onde foi debatido várias questões relacionadas à temática durante um pouco mais de 1 hora. A seleção levou em conta requisitos sobre o tema, tempo de experiência, prática, vivência, entre outros, na tentativa de refinar os dados obtidos para gerar como proposta de implantação das boas práticas de governança de dados aliados ao big data e uma guideline, cumprindo as seguintes etapas:

**1ª Etapa:** Seleção da equipe com conhecimento substancial dos tópicos em discussão com conhecimentos específicos de Governança de Dados e Big Data;

**2ª Etapa:** Disponibilização do Termo de Consentimento Livre e Esclarecido (TCLE), que busca a conscientização do participante de maneira livre e espontânea, informando sobre os objetivos do Grupo Focal, levantamento de dados e informações relevantes sobre os processos;

**3ª Etapa:** Elaboração do roteiro de discussão, dividido em 6 tópicos: Dados, Negócio do Órgão, Liderança, Metas, Tecnologia, Equipe de TI e Analistas de Dados, com 29 questões, conforme Apêndice C.

**4ª Etapa:** Condução da entrevista em um ambiente online e interativo;

**5ª Etapa:** Análise dos resultados.

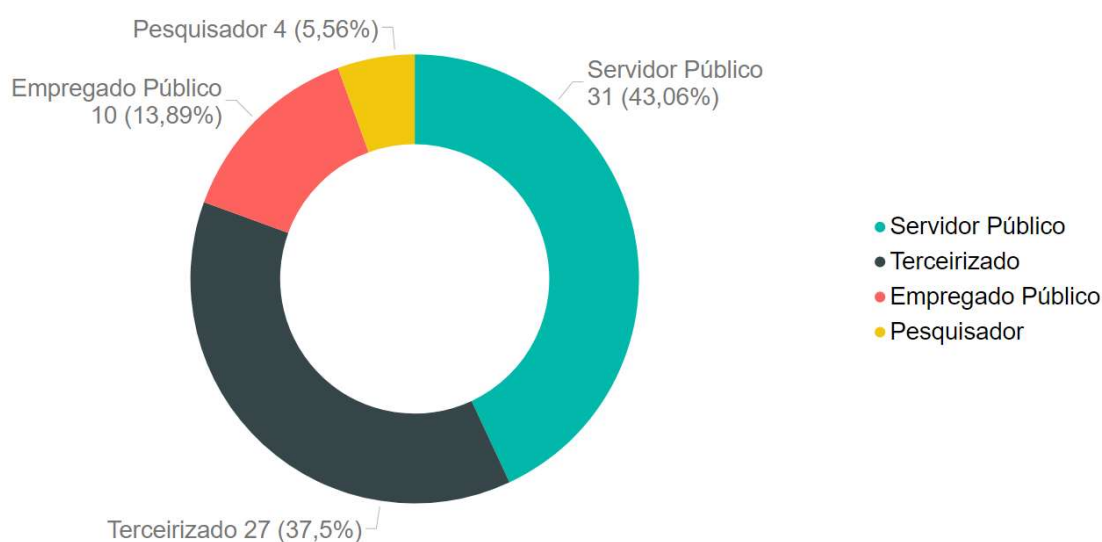
## 4. RESULTADOS E DISCUÇÕES

### 4.1. Análise da aplicação da pesquisa Survey

Inicialmente, cumpre destacar que a pesquisa de campo foi utilizada como ferramenta acessória na busca por respostas às perguntas principal e auxiliares da pesquisa em relato. O intuito da pesquisa era dimensionar o envolvimento dos órgãos públicos federais com as técnicas e diretrizes empregadas ao Big Data.

Entre as questões de 1 a 5 empregadas no questionário em apêndice, se buscou meios de identificação do perfil do público respondente.

Gráfico 1 - Cargo na Administração Pública ocupados pelos entrevistados



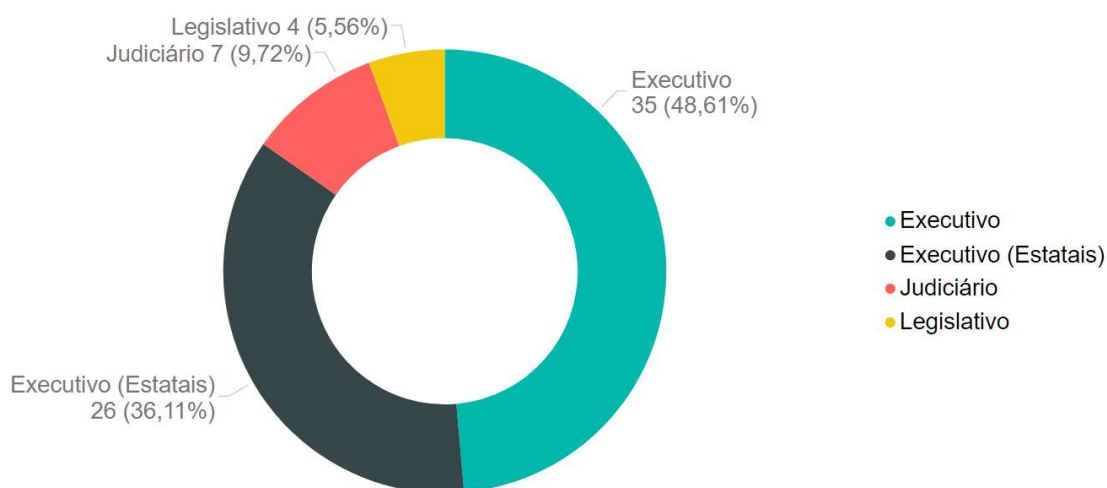
Fonte: próprio autor.

No Gráfico 1 acima, com relação ao cargo ocupado pelos entrevistados na administração pública federal, nota-se que 31 pessoas, ou seja, 43,06% pertencem a classe dos Servidores Públicos. Já 27 pessoas (37,5%) pertencem aos grupos dos profissionais terceirizados. Empregado Público aparece em 3º lugar com 10 pessoas ou 13,89% e como Pesquisadores apenas 4 pessoas ou 5,56%. Logo, a maioria dos entrevistados são servidores e empregados públicos, o que pode representar um público que, embora ligado à área de TI, teria condições de fazer a conexão, em teoria,

dos trabalhos da TI com as atividades institucionais de cada instituição pública representada.

Analisando a segunda questão no Gráfico 02, ligada ao poder que os entrevistados trabalham, a maioria trabalha no poder executivo, representando 35 pessoas ou 48,61%. Enquanto 26 pessoas, ou 36,11%, trabalham no setor Executivo (Estatais), 7 pessoas no Judiciário e 4 pessoas no Legislativo, representando apenas 9,72% e 5,56% respectivamente.

Gráfico 2 – Em qual poder trabalha o entrevistado

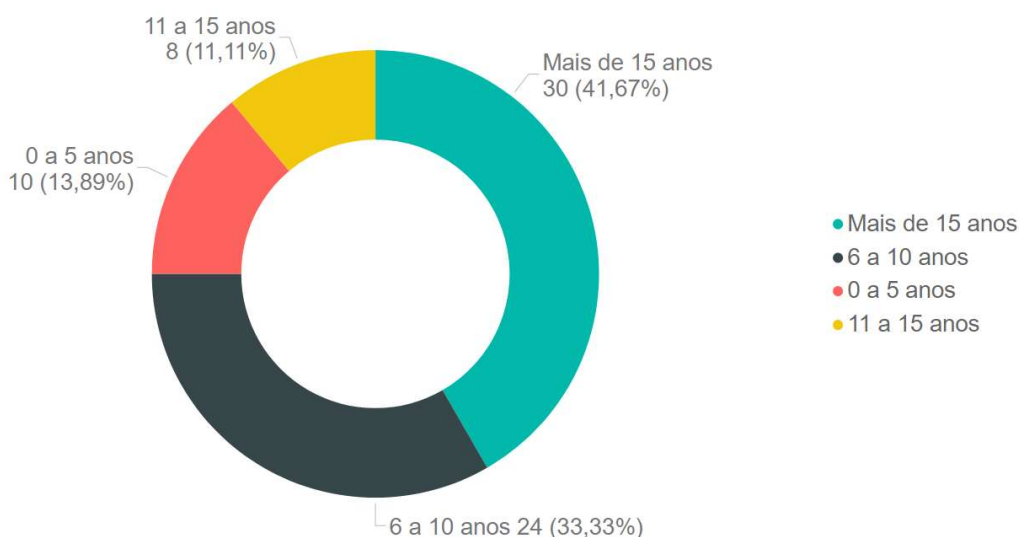


Fonte: próprio autor.

Na terceira questão explorou-se em qual órgão o entrevistado trabalha. A quantidade total foi de 45 órgãos.

A quarta questão, da qual o resultado é compilado no Gráfico 3, explora o tempo de experiência dos entrevistados em relação ao emprego em TI.

Gráfico 3 - Tempo de experiência com T.I do entrevistado.



Fonte: próprio autor.

A maioria (cerca de 52%) possui mais de 10 anos de experiência em seus cargos.

Estas perguntas iniciais trataram de traçar um pouco do perfil dos respondentes e seus locais de trabalho. O perfil indicou que se trata de um conjunto de profissionais com experiência em TI e que trabalham em órgãos majoritariamente do Poder Executivo, portanto, órgãos mais ligados a execução de políticas e serviços públicos, e que, assim, guardariam, a princípio, certas similaridades e resultados de utilização de tecnologias de Big Data.

A partir desta etapa, se inicia os questionamentos que buscam compreender informações com as tecnologias relacionadas com o Big Data. Desta forma, na questão cinco, retrata apenas os estados dos quais os órgãos estão lotados, e na próxima questão, u seja a sexta questão, inicia-se assuntos como a familiarização, definições e termos correlacionados com a aludida tecnologia.

Já na análise da questão 6, verificou qual era a percepção do entrevistado na familiarização ou não com os termos concernentes ao Big Data. Noventa e quatro por cento (94,44%) afirmaram estar familiarizados com a terminologia correlata ao tema.

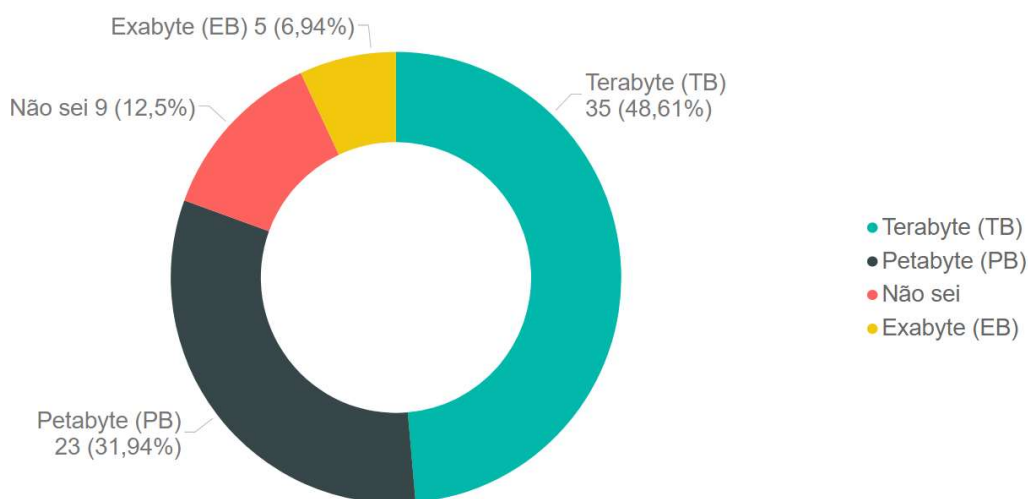
Na sétima questão, inicialmente buscou-se compreender qual era o nível de experiência relacionada ao “assunto” Big Data. Porém mesmo com as repostas, conclui-se que tais questionamentos não significam de fato que o entrevistado possui experiência com o Big Data, até porque esta experiência não pode ser medida de forma direta neste objeto de estudo.

Na oitava questão, buscou-se compreender, na visão do entrevistado, a relação entre a volumetria de dados e o Big Data. A totalidade das respostas indicou que há a relação, indicando que todos os entrevistados estavam cientes que o ambiente de Big Data estaria relacionado com o alto volume de dados. Assim teriam condições de que, mesmo que não concordassem acerca do exato volume de dados que poderia ser considerado “Big Data”, estaria ciente que este volume não pode ser pequeno.

Posteriormente, na nona questão, busca-se saber se no órgão existe ambiente para Big Data. Desta forma, 32% respondem que não e 68% dizem que sim. Esta questão pode ser contrastante com questões anteriores, principalmente com relação à experiência dos entrevistados ligados ao Big Data. Como analisado anteriormente, o fato de se ter um “tempo” trabalhando na área, não significa que o entrevistado possui muita ou pouca experiência na área e também não pode-se dizer de que o fato dos órgãos (no caso 32%) não terem este ambiente, significar que os funcionários não tenham experiência em Big Data.

Na décima questão buscou-se identificar qual era a volumetria de armazenamento de dados do órgão em que trabalha o respondente. A compilação deste resultado está graficamente representada no Gráfico 4.

Gráfico 4 - Volumetria de armazenamento dos órgãos.



Fonte: próprio autor.

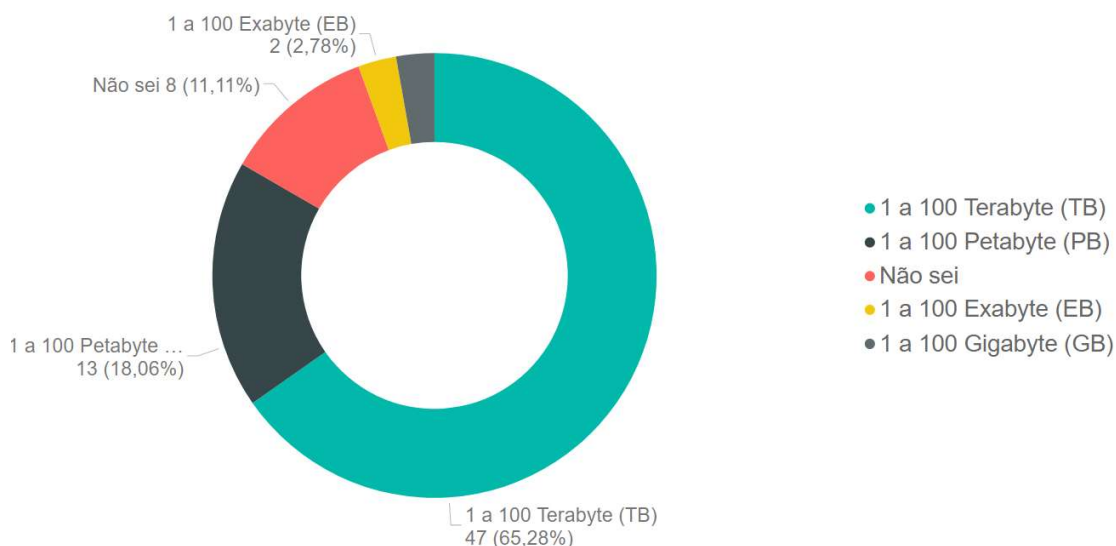
Verifica-se que boa parte dos órgãos trabalha com Terabytes de informação, 48,61%, na visão dos entrevistados. Outro grupo indicou volumes na casa de Petabytes de informações, representando 31,94%. Doze por cento (12,5%) não souberam responder e 6,94% trabalham com Exabytes de informações. Terabytes, contudo, pode não representar, necessariamente, um ambiente de Big Data.

Assim, a não ser que as tecnologias utilizadas fossem indiscutivelmente relacionadas com o Big Data, poderia haver questionamento acerca das respostas anteriores quanto à existência de ambientes de Big Data, pois com base na volumetria de Terabytes não há segurança na afirmação da existência de tal ambiente.

Continuando, na décima primeira questão perguntou-se qual é a projeção do crescimento do volume de dados para o ano de 2020 (Gráfico 5).



Gráfico 5 - Projeção de crescimento do banco de dados.

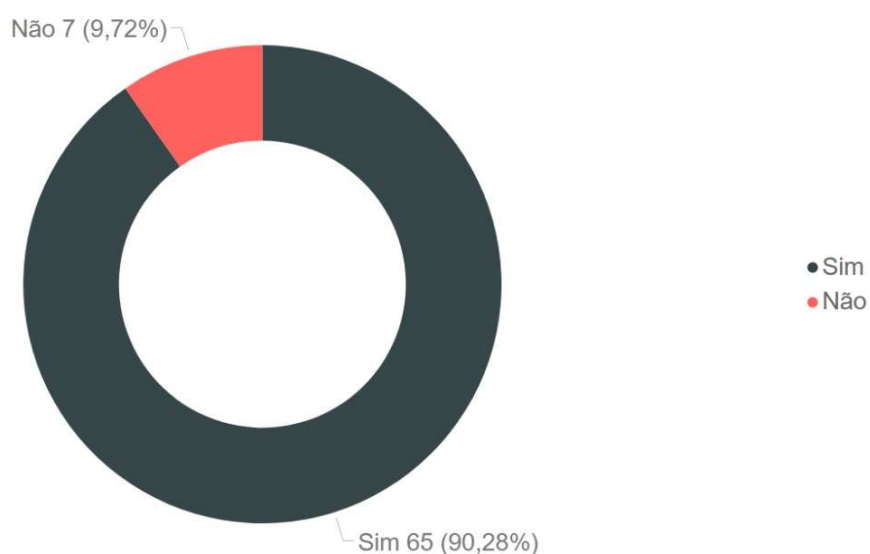


Fonte: próprio autor.

Sessenta e cinco por cento 65,28% disseram que era esperado entre 1 a 100 Terabytes de volume de dados, 18,06% entre 1 a 100 Petabytes, menos de 2,78% esperam de 1 a 100 Gigabyte e Exabytes respectivamente e 11,11% não souberam responder. Este resultado se encontra compilado no Gráfico 5.

Já na décima segunda questão questionou-se se com o crescimento esperado, o órgão onde trabalha o respondente, passaria a ter um ambiente de Big Data. O resultado está compilado no Gráfico 6.

Gráfico 6 - Possui projeto para ambiente próprio de Big Data?



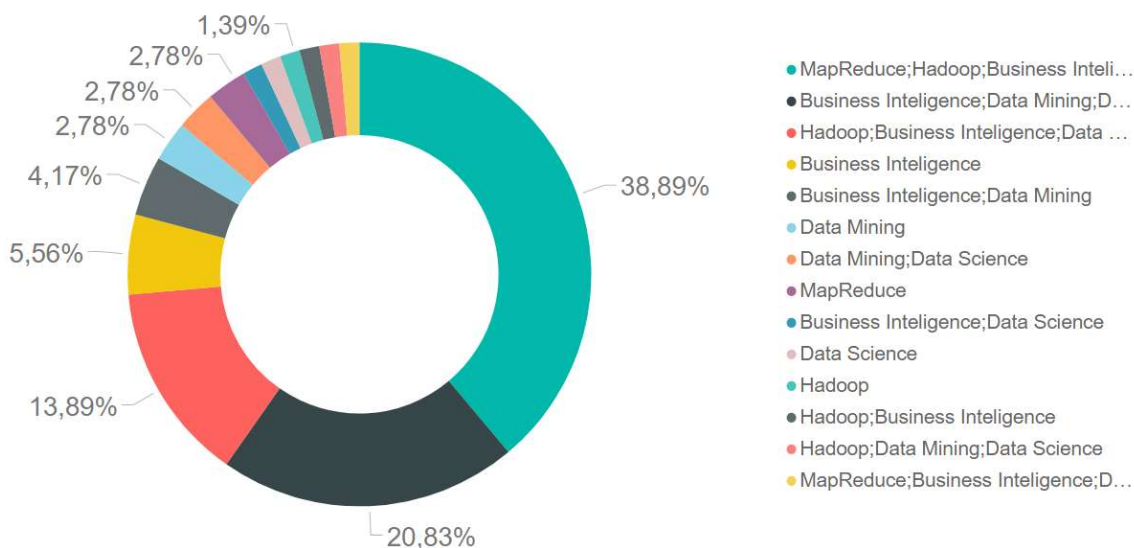
Fonte: próprio autor.

Mais de noventa por cento (90,28%) afirmaram que possuem projetos para aderirem ao próprio ambiente de Big Data e 9,72% não possuem.

Ou seja, se confirmou que a maioria dos profissionais respondentes, de fato, compreende que seus volumes de dados na casa de Terabytes está suficiente para caracterizar o seu ambiente de Big Data, e que portanto, para estes, a influência das tecnologias componentes do Big Data para a definição da existência de um ambiente é mais importante que volumes que cheguem a Petabytes, por exemplo.

Na questão 13 (Gráfico 7) e 14 (Gráfico 8) buscou-se verificar se os usuários estariam familiarizados com algumas tecnologias relacionadas com o Big Data, e se sabiam diferenciá-las de outras também comuns nestes ambientes.

Gráfico 7 - Diferenciação entre os termos distintos de Big Data.



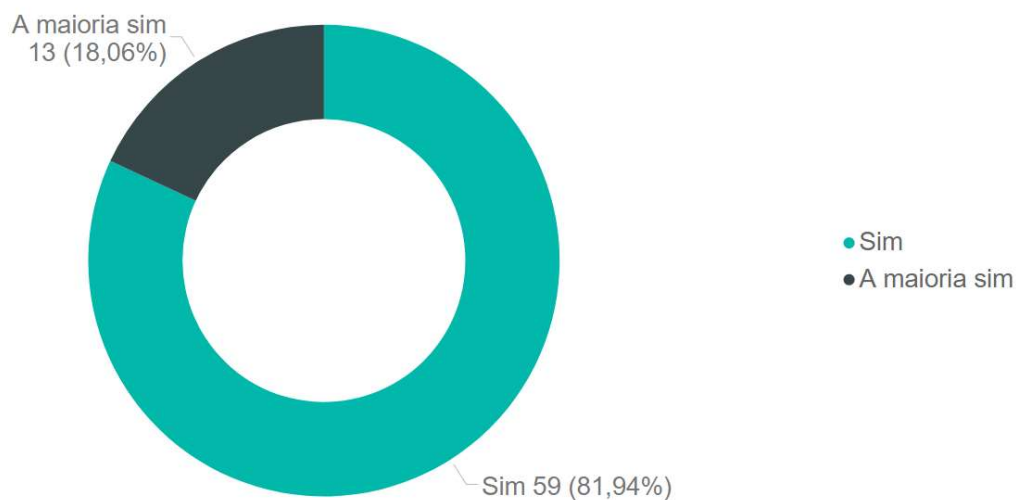
Fonte: próprio autor.

Os três maiores grupos de respondentes a esta questão, indicaram que a maior parte dos respondentes sabia diferir pelo menos três dos conceitos. Estes três maiores grupos compuseram cerca de 72% das respostas. O resultado, ainda, demonstrou que mais de um terço deles sabia diferir todos os conceitos. O gráfico 7 acima, compila os resultados.

O resultado demonstra que ao menos uma parte dos que afirmaram existir um ambiente de Big Data em suas repartições conhecia inclusive acerca de algumas das tecnologias do Big Data, e que, portanto, era um fato que a composição deste ambiente estaria ligada às tecnologias, além do volume de dados propriamente dito, o que não apenas corrobora que eles conheciam o conceito, mas que de alguma forma o aplicavam aos seus próprios ambientes profissionais.

A décima quarta questão serviu de complemento à décima terceira, verificando se os respondentes sabiam acerca da profundidade de relação entre os conceitos apresentados e o ambiente de Big Data.

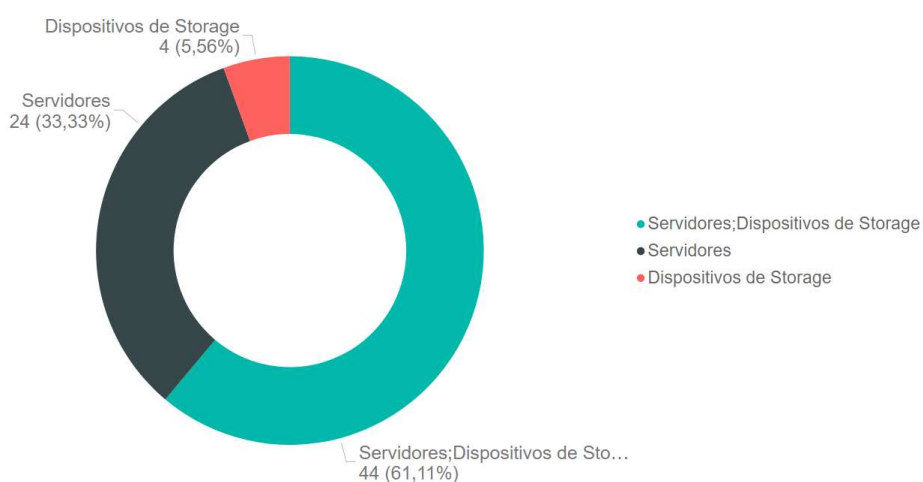
Gráfico 8 - Considera que todos os termos acima possuem forte relação com o tema Big Data?



Fonte: próprio autor.

Oitenta e um por cento (81,94%) tinha em mente que os conceitos/tecnologias estariam fortemente relacionados com o Big Data, conforme demonstrado no Gráfico 8 acima. O resultado que se encontra no Gráfico 9, relativo a décima quinta questão, pode ser analisado na seguinte forma:

Gráfico 9 - Tipos de armazenamento de dados do órgão.

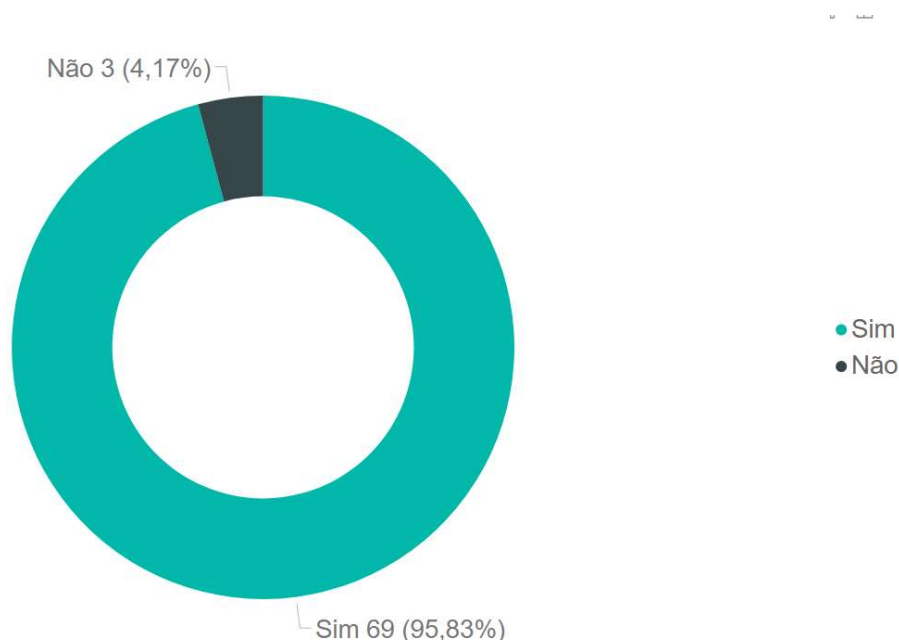


Fonte: próprio autor.

A décima quinta questão referia-se à quais os tipos de armazenamento de dados do órgão. Sessenta e um por cento (61,11%) disseram Servidores Dispositivos de Storage; 33,33% marcaram Servidores; e 5,56% responderam Dispositivos de Storage. Este resultado se encontra compilado no Gráfico 9.

A questão a seguir, décima sexta, buscava conhecer sobre a infraestrutura de armazenamento de dados estruturados e não estruturados. O resultado está apresentado no Gráfico 10.

Gráficos 10 – Possui dados não estruturados



Fonte: próprio autor.

É possível notar que noventa e cinco por cento (95,83%) disseram que sim, havia dados não estruturados e 4,17% disseram que não, ou seja, apenas 4,17% destes dados relativos aos órgãos pesquisados são estruturados.

Já na décima sétima questão, verificou-se sobre os órgãos possuir acesso a um grande volume de dados (tanto estruturados como não estruturados). Deve-se ressaltar que em um ambiente de Big Data, deve ser considerado os 5 Vs (Volume, Velocidade, Variedade, Veracidade e Valor), porém especificadamente nesta questão,

busca-se a compreensão sobre o acesso dos órgãos a este “volume” de dados “estruturados e não estruturados”.

Este acesso pode significar que, embora o órgão não possua um alto volume de dados, ele tem acesso a coletar informações de grandes volumes, muitas vezes não estruturados, demandando que tenha ferramental e ambiente necessário para o processamento que caracteriza o Big Data.

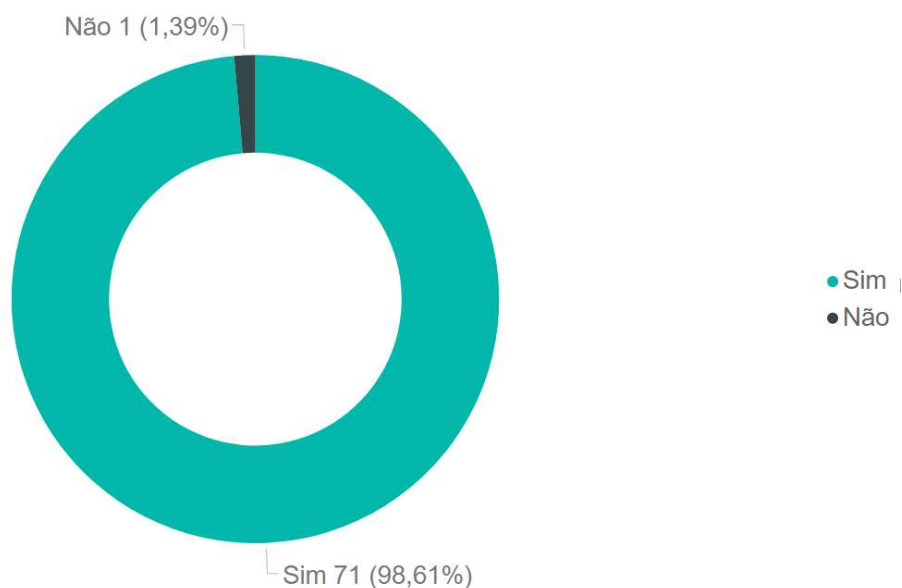
Até esta questão, em resumo, o cenário indicava que os respondentes tinham noção suficiente do que era o Big Data, sabiam um pouco sobre as tecnologias relacionadas, seus órgãos tinham pessoa com algum nível de conhecimento na tecnologia, tinham acesso a um grande volume de dados e tinham ambientes e projetos de Big Data em seus ambientes profissionais.

Esta coleta de informações que demandou cerca de metade das perguntas do questionário objetivou, em suma, aferir se as demais respostas do questionário, aquelas que se alinham com as perguntas principais e secundárias da presente dissertação poderiam de fato contribuir para o entendimento foco do trabalho.

Verificou-se que sim, havia ambientes de Big Data na Administração Pública Brasileira, segundo a amostra de respostas do questionário, que está contando com profissionais que tem as noções mínimas da tecnologia, e que estão, em sua maioria, com compromissos profissionais de longa data com suas respectivas repartições. Com este cenário, foi possível, então, na continuidade da análise das respostas do questionário averiguar aspectos do alinhamento da TI de Big Data descrita com os negócios das repartições públicas representadas, permitindo a aferição, ao menos dentro da amostra obtida, quanto aos questionamentos do trabalho em tela.

Na décima oitava questão (Gráfico 11), foi perguntada aos entrevistados se existiria algum problema de negócio que poderia ser resolvido com as tecnologias voltadas para Big Data.

Gráficos 11 - Existe algum problema de negócio que poderia ser resolvido com Big Data?

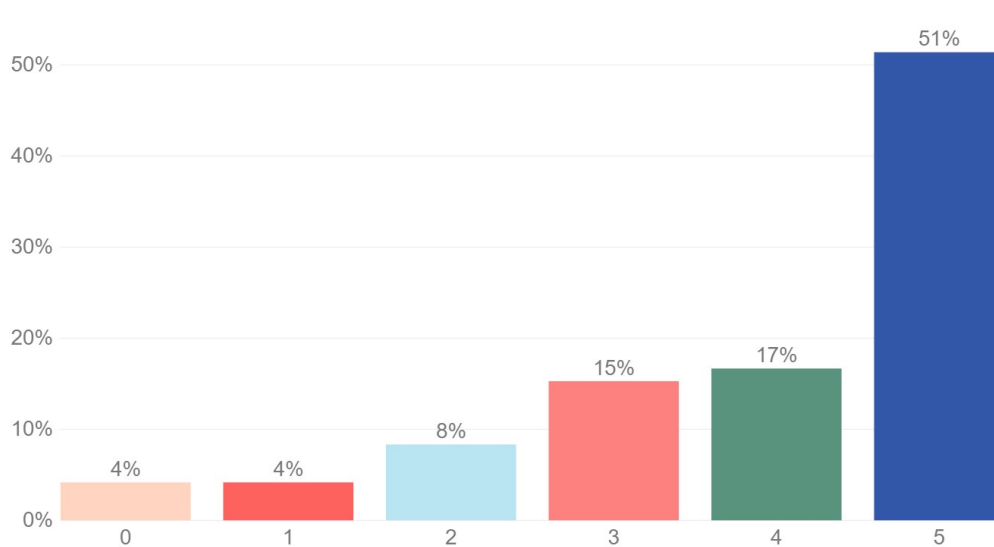


Fonte: próprio autor.

Noventa e oito por cento (98,61%) disseram que sim, que tinham problemas que poderiam ser resolvidos. Um por cento (1,39%) disse que não. Os entrevistados, portanto, sabiam sobre esses problemas, demonstrando uma janela de oportunidades para aplicações de Big Data, iniciando as respostas aos questionamentos centrais da pesquisa.

Na décima nona questão, buscou-se saber se havia prospecção de implantação de Big Data nos órgãos que ainda não tinham a aludida estrutura; o resultado está compilado no Gráfico 12.

Gráficos 12 - Perspectiva para implantação de Big Data nos órgãos.

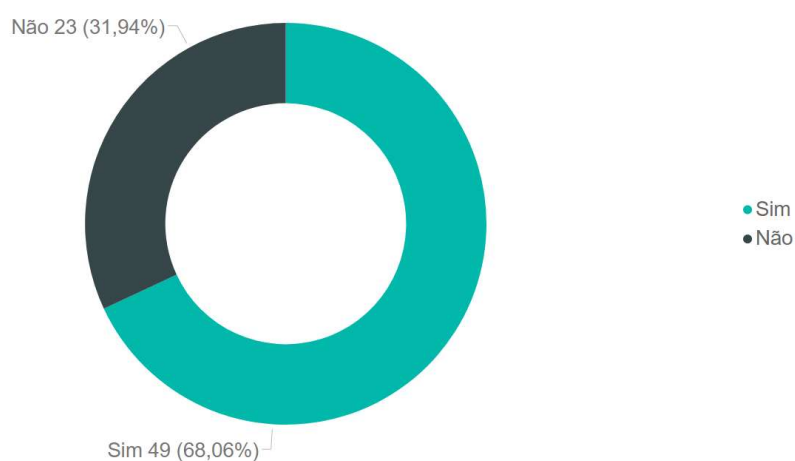


Fonte: próprio autor.

Esta pergunta demonstra que além de muitos órgãos já possuírem o Big Data, estão com novos projetos de utilização.

Isso ficou ainda mais claro quando se analisou a resposta da pergunta seguinte, na qual referia se havia projetos de Big Data de uma maneira mais direta, sendo que houve resposta positiva de 68,06% dos entrevistados.

Gráficos 13 - O órgão já possui projetos de Big Data?



Fonte: próprio autor.



No entanto, o que estaria a motivar estes projetos? Essa seria a pergunta mais adequada no sentido de apoiar os questionamentos principais da dissertação. A que oportunidades de negócio o Big Data estaria a ser uma resposta ou solução?

A pergunta 21 tratou dessa questão. Nas respostas, compiladas no Gráfico 14, se verificou que o maior motivador era a missão institucional do órgão representado.

Gráficos 14 - O que tem motivado o uso de Big Data no seu órgão público?



Fonte: próprio autor.

Isso demonstrou alinhamento com o fato descrito nos capítulos anteriores, ou seja, o Big Data precisa estar alinhado ao atendimento da missão institucional pra que seus projetos sejam utilizáveis, ou seja, o maior motivador, a maior oportunidade de implantação do Big Data deve depender de forma direta desta característica.

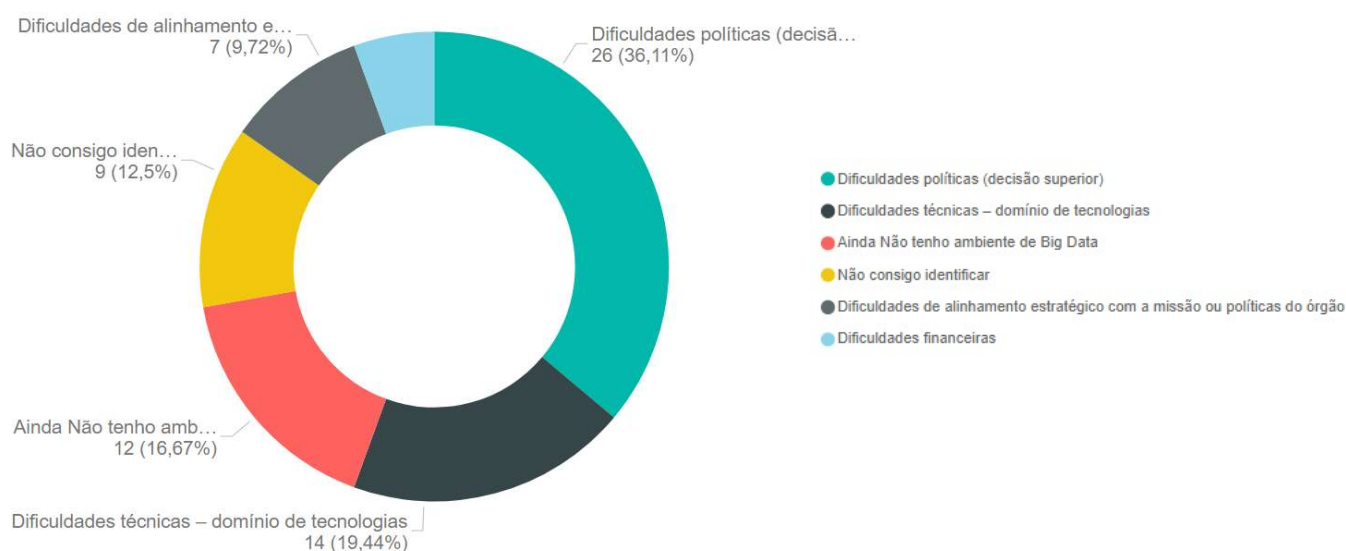
Um grupo respondeu que a motivação tinha relação com avanços tecnológicos. O fato é preocupante, pois demonstra um grupo de pessoas que não tem encontrado no Big Data uma solução de problemas públicos, mas o inverso, precisa de motivações no órgão para justificar investimentos em Big Data.

Um grupo ainda relacionou os projetos com a eficiência de gastos públicos. No entanto, se tratou de um grupo representativo. Há ainda de se destacar que um grupo considerável relaciona os projetos de Big Data com questões de transparência. Dessa

forma, esse aspecto alinha as respostas com as expectativas iniciais da pesquisa em relato.

Quanto às dificuldades encontradas, objetos da pergunta vigésima segunda, apresenta-se o Gráfico 15.

Gráficos 15 - Dificuldades encontradas na implementação do ambiente de Big Data.

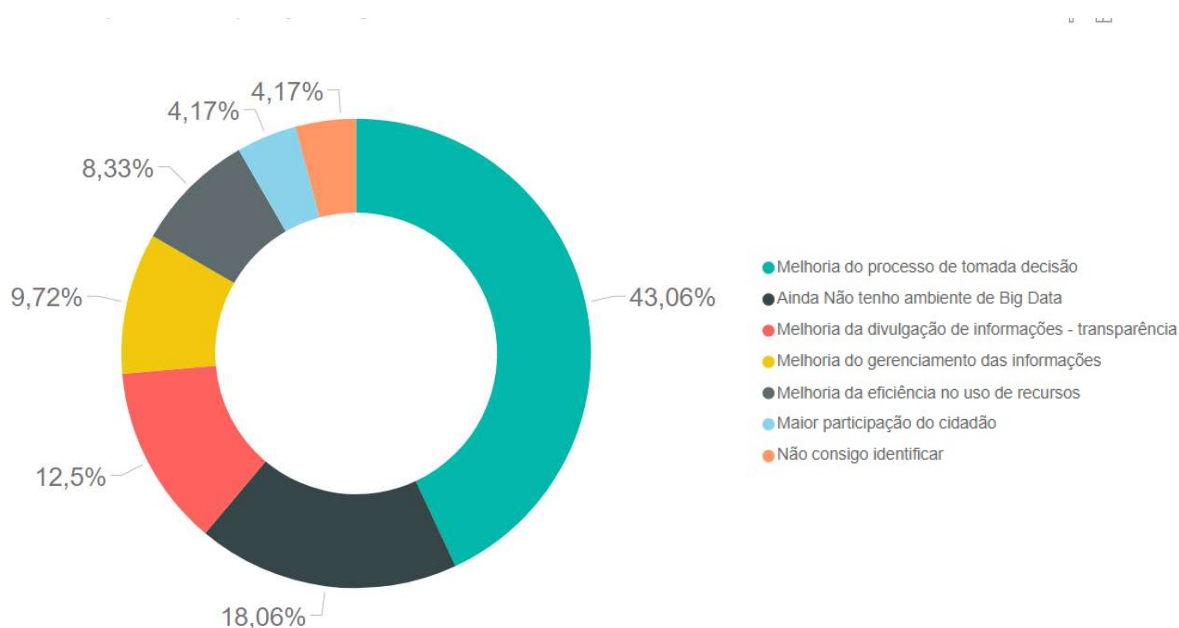


Fonte: próprio autor.

Os grupos de destaque elencados como dificuldades na implantação de projetos de Big Data foram os das Dificuldades políticas e Dificuldades técnicas. Há, portanto, um enorme desafio ligado a estas duas questões. As dificuldades políticas podem estar relacionadas com questões burocráticas e políticas de fato. Determinados projetos da Administração Pública Brasileira sofrem de descontinuidades derivadas das trocas políticas frequentes nos escalões superiores dos órgãos públicos. Estas respostas também se mostraram muito úteis quando da revisão das perguntas norteadoras da pesquisa.

A pergunta seguinte buscou aferir quais os principais resultados, impactos com a implantação do Big Data, acompanhada de uma pergunta mais relacionada com questões de eficiência no gasto de recursos públicos (Gráfico 16).

Gráficos 16 - Impactos com a implantação de Big Data.

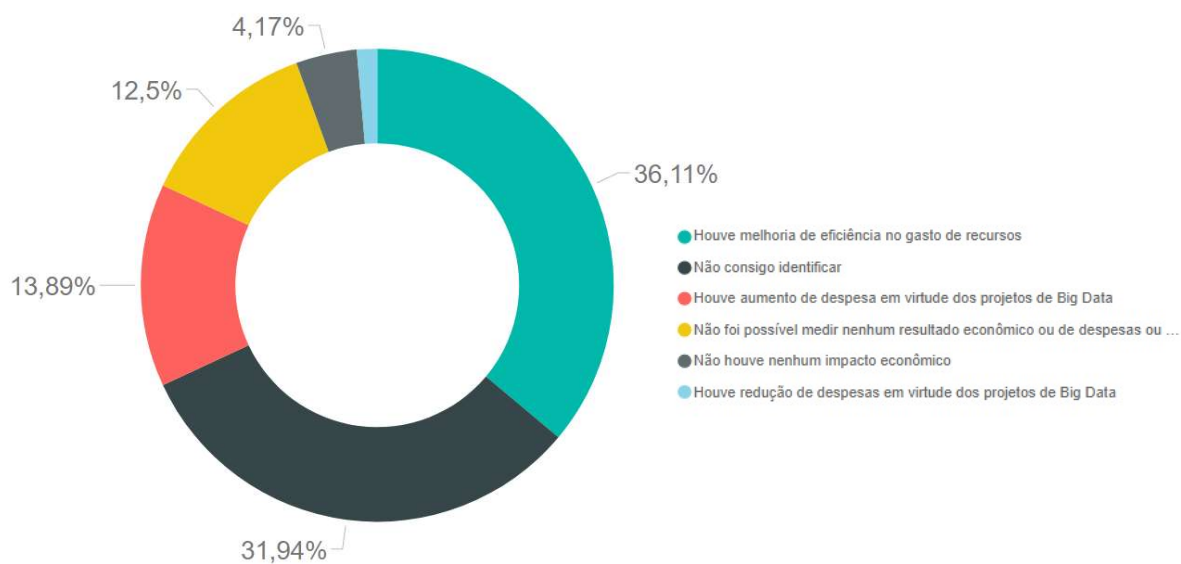


Fonte: próprio autor.

As respostas a estas questões destacaram que o Big Data tem sido um promotor de melhoria de tomada de decisões e que tem sido um ótimo contribuinte quanto a questões de transparência.

Economicamente, parte dos respondentes informou que houve melhoria quanto à eficiência de gastos públicos, embora uma parte considerável não tenha conseguido identificar os impactos econômicos. Esperava-se que uma parte maior dos entrevistados conseguisse identificar melhorias de eficiência econômica.

Gráficos 17 - Os impactos econômicos da adoção de Big Data nos órgãos.

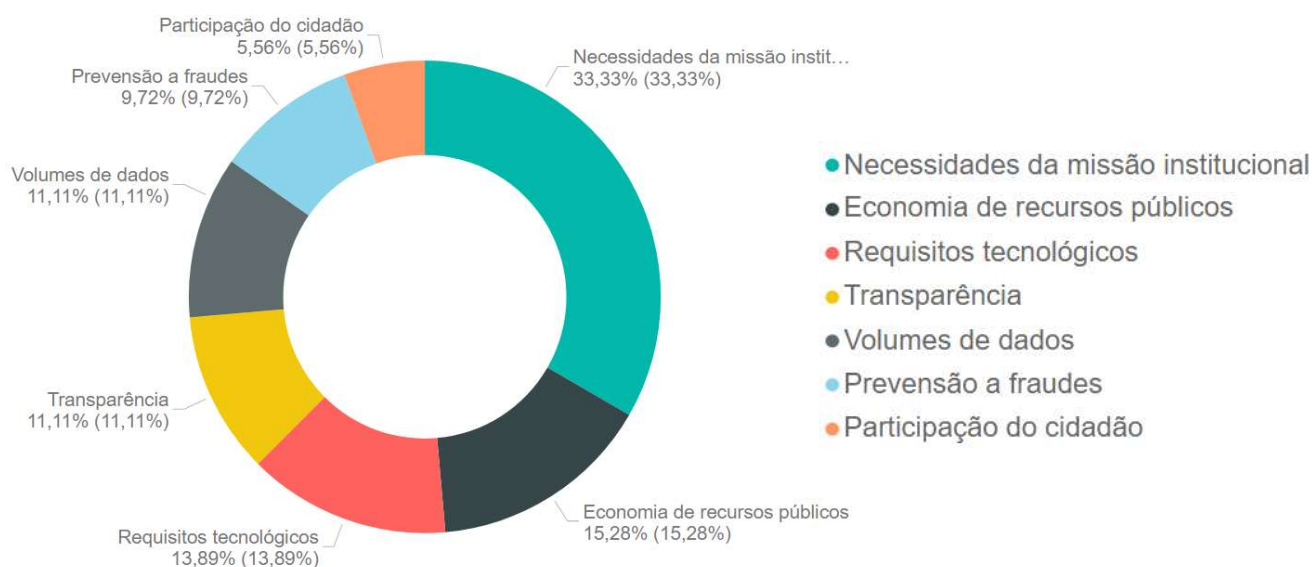


Fonte: próprio autor.

Um resultado considerado interessante foi o de que parte do grupo identificou um aumento de gastos. Não ficou claro, contudo, se este aumento estaria ligado apenas ao setor de TI.

A pergunta seguinte questionava os usuários acerca das oportunidades de utilização do Big Data, ou seja, não apenas quanto a projetos que já utilizam, mas quanto a motivadores de novos projetos de Big Data no setor público (Gráfico 18).

Gráficos 18 - As oportunidades de implementação do Big Data nos órgãos públicos.

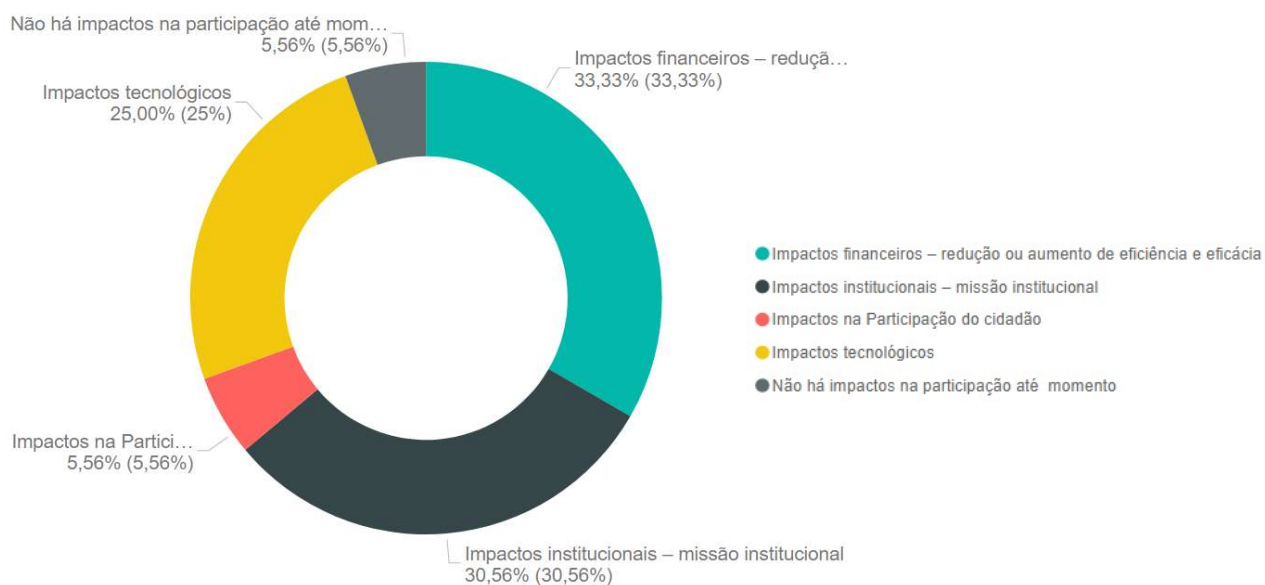


Fonte: próprio autor.

As respostas indicaram como os dois maiores motivadores as necessidades relacionadas com a missão institucional e a economia de recursos públicos. Chama a atenção, contudo, o fato de que poucos relacionam as oportunidades novas de utilização do Big Data com a maior participação do cidadão nos sistemas destes órgãos públicos. Os resultados são apresentados no Gráfico 18 (acima).

A vigésima sexta questão buscou aferir quais os impactos que os projetos de Big Data podem trazer às instituições representadas, conforme demonstradas no Gráfico 19 abaixo.

Gráficos 19 - Os impactos que práticas de Big Data podem/poderiam trazer ao desenvolvimento de soluções nos órgãos públicos.

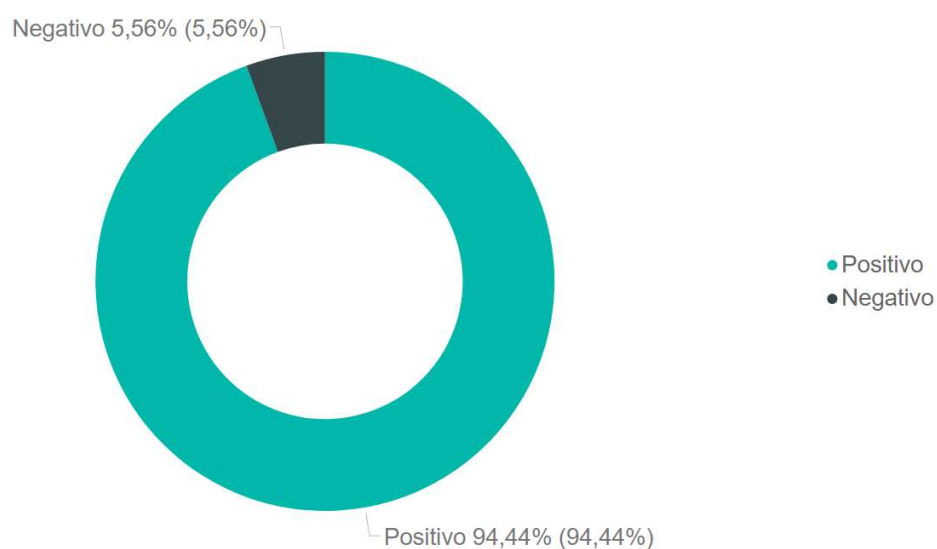


Fonte: próprio autor.

Três grupos de resposta se destacam: Impactos financeiros (eficiência no gasto público), Impactos na missão institucional (o que também se traduz em eficiência e eficácia) e impactos tecnológicos (como a TI impacta e é impactada com os aludidos projetos). Os resultados estão compilados no Gráfico 19 (acima).

Este questionamento prossegue na questão 27, representada pelo gráfico 20 abaixo, onde verifica-se se os impactos sobre os entrevistados são mais positivos ou negativos.

Gráficos 20 - Os impactos citados na questão anterior são mais positivos que negativos nos órgãos públicos?

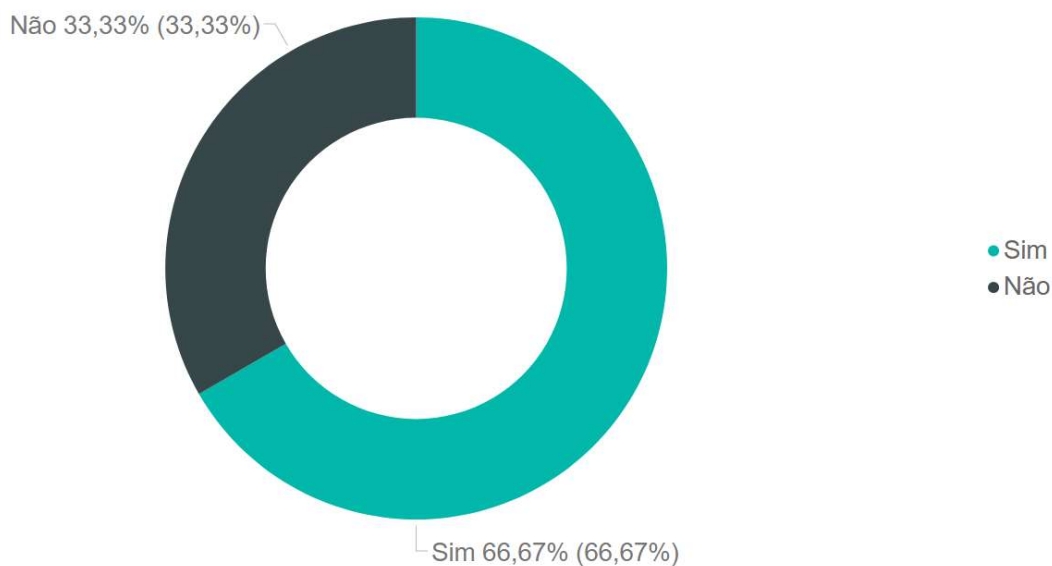


Fonte: próprio autor.

A questão, inicial, pode parecer muito óbvia. No entanto, como exemplo, se tomarmos que um grupo se referiu a impactos tecnológicos, estes poderiam ser negativos caso houvesse a necessidade de implantação de itens de tecnologia para os quais uma instituição não estivesse preparada para implementar. Igual, uma implantação com condução incorreta poderia afetar o cumprimento da missão institucional. No entanto, a maioria dos entrevistados (94,44%) identificou que os impactos seriam positivos.

A pergunta 28 também estava relacionada com estes impactos. Perguntou-se se os impactos estariam ligados a casos ou situações reais ou conhecidas pelos entrevistados (Gráfico 21).

Gráficos 21 - A sua resposta à questão anterior envolveu impactos existentes e comprovados (Sim) ou impactos percebidos pelo conhecimento acerca de Big Data conhecido pelo respondente do questionário (Não).



Fonte: próprio autor.

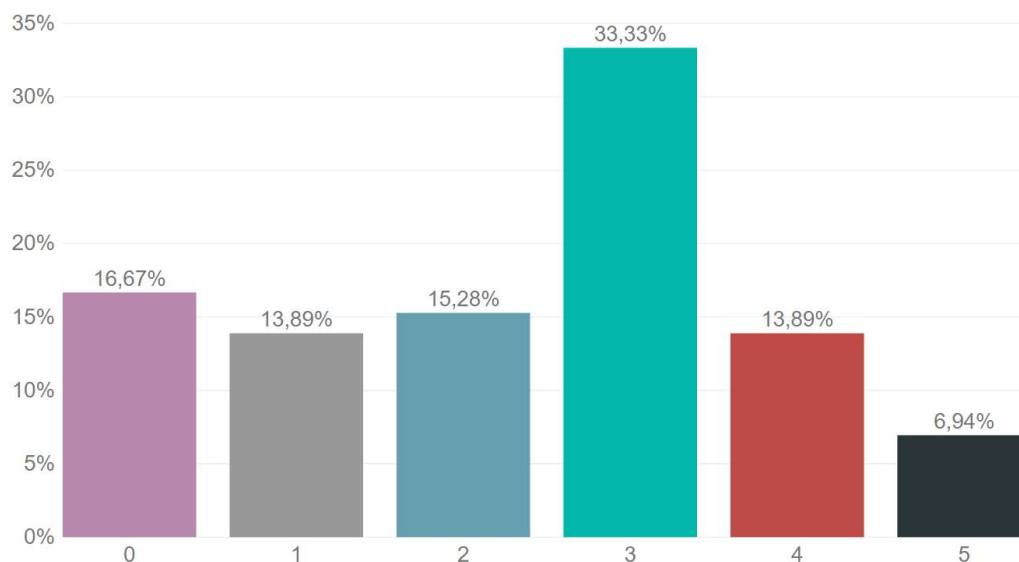
Sessenta e seis por cento (66,67%) afirmaram que se trata de casos reais conhecidos, ou seja, relacionam que a sua medição de impactos que podem ocorrer com projetos de Big Data é positiva de acordo com seu conhecimento de experiências anteriores.

Da vigésima nona questão à trigésima terceira questão se buscou conhecer um pouco o quadro de distribuição de conhecimento e de presença de profissionais de ciência de dados, bem como algumas informações de utilização de tecnologia de nuvem.

A distribuição de conhecimento se mostrou com uma intensidade média, e os resultados estão compilados nos Gráficos 22 e 23, relacionados com a distribuição de conhecimento entre os profissionais de TI e entre o setor de TI e outros setores das instituições.

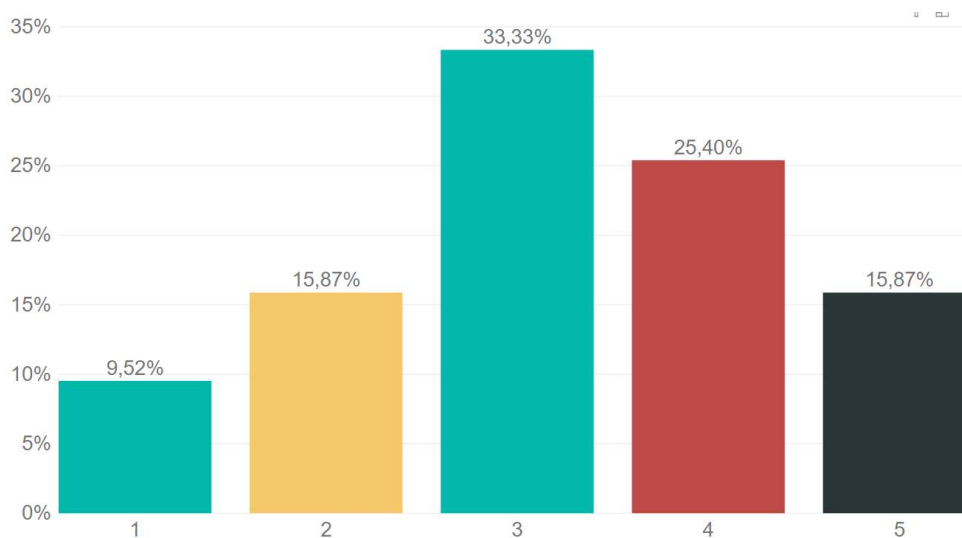


Gráficos 22 - O órgão compartilha conhecimentos e aplicações de Big Data entre todos os setores do órgão?



Fonte: Próprio autor

Gráficos 23 - A equipe de TI e profissionais de dados compartilha conhecimentos entre todos os setores da organização?

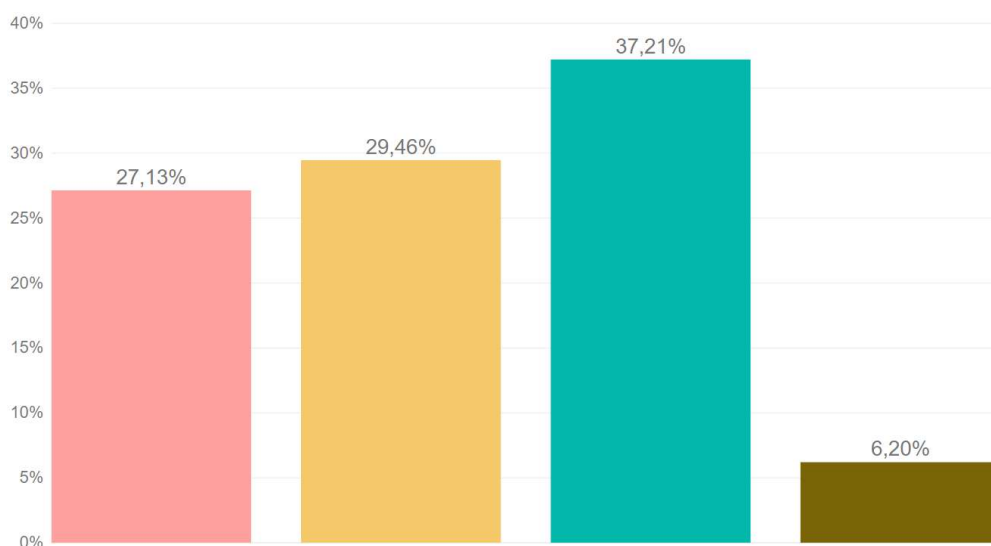


Fonte: próprio autor.

A questão 31 buscou verificar o uso de repositórios e serviços em nuvem para processamento dos dados (haveria o uso em que intensidade). A intensidade de uso

destes repositórios foi considerada média, o que remete à ideia de que parte pode estar ocorrendo dentro das infraestruturas das instituições representadas. O resultado está compilado no Gráfico 24.

Gráficos 24 - No ecossistema de Big Data é explorado ou adotado serviços de nuvem para processar ou conduzir o Analytics?

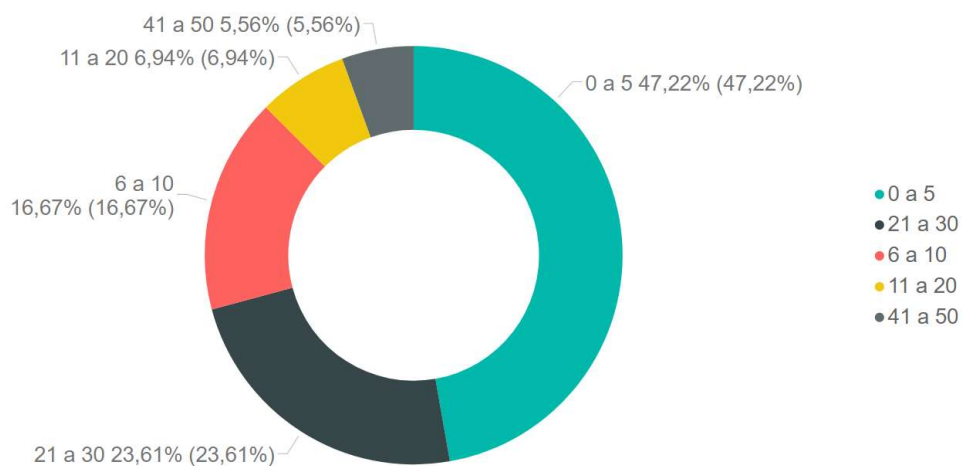


Fonte: próprio autor.

Assim, as duas perguntas finais buscaram aferir a quantidade de servidores cientistas de dados, tanto servidores públicos concursados (pergunta 32) quanto terceirizados (pergunta 33). Os resultados estão compilados nos Gráficos 25 e 26, logo abaixo.

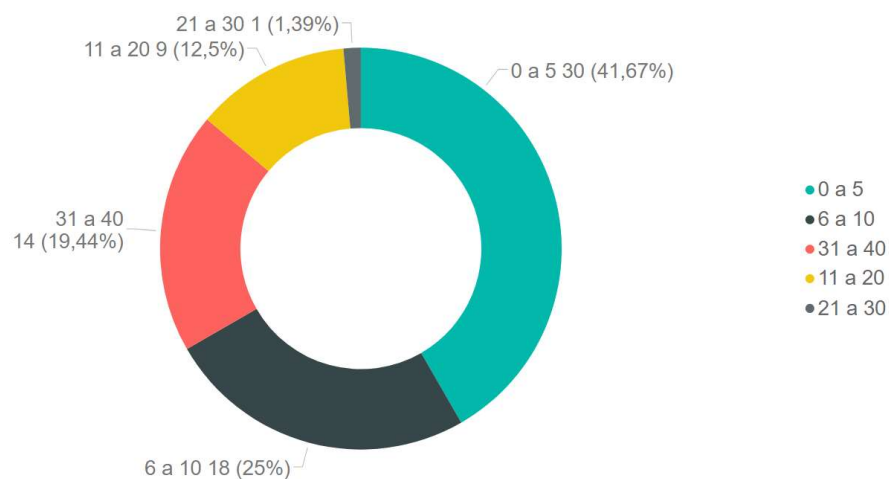
No entanto, alguns poucos respondentes parecem não terem entendido as questões corretamente, uma vez que seria improvável, dado o tamanho do corpo de servidores em TI nos órgãos públicos, que em algum órgão pudesse haver quantidade de servidores dedicados à ciência de dados maior que 30 ou 40 pessoas.

Gráficos 25 - Quantos profissionais de cientistas de dados o órgão possui servidores públicos?



Fonte: próprio autor.

Gráficos 26 - Quantos profissionais de cientistas de dados o órgão possui terceirizado?



Fonte: próprio autor.

Realizando um apanhado geral sobre o questionário aplicado, sabe-se que o intuito é verificação e a empregabilidade das boas práticas nos órgãos públicos federais. Diante tal informação (objetivo), ressalta-se que uma fatia dos entrevistados é oriunda do setor público, ou seja, estão inseridos no cenário de desejo. Destes profissionais envolvidos, boa parte dos mesmos trabalham no setor executivo em seus órgãos públicos e também possuem experiência quando relacionado suas tarefas e as aplicações sobre tecnologia da informação.

Apesar do questionário ter sido empregado através do Google Forms, ou seja, disponibilizado em território nacional através do portal da Google, boa parte dos entrevistados são do Distrito Federal, mas há registros de participação de outros estados como Acre, Piauí, Paraná, entre outros. Apesar da tecnologia avançar em passos largos e muitas vezes os trabalhadores não conseguem seguir tais avanços, o questionário denota um ponto interessante sobre a experiência dos entrevistados em relação as diretrizes e boas práticas relacionadas ao Big Data, ou seja, há uma aceitação e conhecimento sobre os termos empregados neste objeto de estudo.

Muitas vezes, nos deparamos com estruturas ruins em órgãos públicos, ou até mesmo a falta de recursos para investimentos em tecnologia, matéria humano, capacitação, etc, porém para este objeto de estudo, denota-se que boa parte dos órgãos públicos que participaram da entrevista possuem ambiente próprio para a realização e empregabilidade de diretrizes para Big Data, além de investimentos próprios para volumes de dados, que perante o questionário tendem a crescer cada vez mais.

Ouro ponto que merece destaque, é a questão sobre a diferenciação entre os termos distintos relacionados ao Big Data, que confundem até mesmo profissionais ligados a área. Neste caso, boa parte dos entrevistados, ou seja, 38% dizem ter conhecimento sobre tais termos, sabem suas diferenças e real empregabilidade, mostrando que possuem conhecimento suficiente para empregar as diretrizes e boas práticas relacionadas a temática.

Pontos explorados como problemas relacionados a negócios que poderiam ser resolvidos com a aplicação de Big Data também foram bastante questionados, e boa parte ou na verdade a maioria acreditam que a prática diária das diretrizes de Big Data podem auxiliar e muito nas decisões e nas tarefas diárias dos órgãos que participaram. Para os órgãos que não possuíam ambiente próprio para direcionar requisitos para as aplicações de Big Data, vários deles possuem projetos que viabilizam a empregabilidade e implantação de Big Data nestes órgãos, facilitando assim análises e trabalhos futuros.

Apesar das dificuldades encontradas na aplicação, conforme demonstrado nas análises anteriores, vários órgãos encontram-se motivados para realizar modificações estruturais para implantação de Big Data, além de demonstrarem interesse em capacitação de profissionais e alocação de recursos para o crescimento da volumetria de dados.

No geral, é percebido um esforço dos órgãos públicos que ainda não possuem um ambiente próprio para Big Data. É válido dizer que tais órgãos estão comprometidos com projetos futuros, a implementar as rotinas e diretrizes nas instituições públicas, fortalecendo as diretrizes, aumentando leque de possibilidades, já que, conforme explorado na revisão da literatura deste objeto de estudo, a importância de Big Data para o setor público federal é almejada pelos governos, não só aqui no Brasil, mas de uma forma globalizada.

## **4.2. Análise da aplicação da pesquisa de Grupo Focal**

Assim, para análise e interpretação dos dados coletados, cumpre destacar que foi utilizada a técnica de pesquisa qualitativa, de Grupo Focal para obter dados a partir de reuniões em grupo e compreender a percepção com as técnicas e diretrizes empregadas na Governança de Dados e ao Big Data.

### **4.2.1 Perfil dos participantes**

Os participantes foram selecionados a partir de suas experiência com iniciativas de Big Data e Governança de Dados e suas posições estratégicas nos Órgãos Públicos.

Compuseram o grupo focal 3 Engenheiros de Dados 4 Cientistas de Dados e a média de experiência com Governança de Dados e Big Data de 4 anos. Ainda, 4 participantes com Mestrado e 3 com especialização.

### **4.2.2 Avaliação de Prontidão para Big Data – Grupo Focal**

Para uma melhor análise dos resultados, foi realizada uma abordagem qualitativa para estabelecer o Ranking Médio (RM) para o questionário que utilizou Escala Likert de 5 pontos para mensurar o grau de concordância ou discordância das questões avaliadas, por meio da obtenção do RM da pontuação atribuída às respostas, relacionando à frequência das respostas dos respondentes que tal atribuição. A escala de intensidades do tipo Likert conta com cinco pontos que foram distribuídos da seguinte forma: 1. Discordo totalmente, 2. Discordo um pouco, 3. Não concordo nem discordo, 4. Concordo um pouco, 5. Concordo totalmente. Assim, os valores menores que 3 são considerados como discordantes e, maiores que 3, como concordantes, considerando uma escala de 5 pontos. O valor 3 seria considerado “indiferente” ou “sem opinião”.

Para o cálculo do RM utilizou-se o método de análise de escala do tipo Likert apresentado por (MALHORTA, 2001 e OLIVEIRA, 2005)

Tabela 2 – Avaliação de Prontidão para Big Data – Grupo Focal.

QUESTÕES		FREQUENCIA DE SUJEITOS					
		1	2	3	4	5	RM
<b>DADOS</b>							
1	No órgão, temos acesso a grandes volumes de dados não estruturados ou os dados chegam em tempo real e em grande velocidade para análise.	2			2	3	
2	No órgão, agrupamos dados de várias fontes internas em um Data Warehouse ou Datamart para facilitar o acesso.	2			4	1	
3	No órgão, mantemos definições e padrões consistentes nos dados que usamos para análise.		5	1	1		
4	No órgão, usuários, tomadores de decisão e desenvolvedores de soluções para áreas do órgão confiam na qualidade de nossos dados.		2	2	3		
<b>NEGÓCIO DO ÓRGÃO</b>							
5	Empregamos institucionalmente uma combinação de abordagem de Big Data e Governança de Dados para atingir as metas do órgão.		3	1	2	1	
6	A nossa gestão assegura que as unidades de negócio e as áreas do órgão trabalhem para estabelecer as prioridades da organização relativas ao Big Data e Governança de Dados.	2	5				
7	Capacitamos nossa equipe de TI e profissionais de análise de dados para fomentar a aprendizagem e compartilhar conhecimentos entre todos os setores da organização.	1	2	2	2		
8	As nossas iniciativas e infraestrutura de Big Data e Governança de Dados recebem aportes e recursos adequados para desenvolver as competências e recursos dos quais precisamos.		1	2	4		
9	Colaboramos com parceiros de áreas, com Órgãos para os quais prestamos suporte e outros integrantes do nosso ecossistema de negócios para compartilhar conteúdo e aplicações de Big Data.	4	2			1	
<b>LIDERANÇA</b>							
10	Os dirigentes e líderes que estão na instituição há mais tempo ponderam regularmente sobre oportunidades que o Big Data e a Governança de Dados oferecem ao nosso negócio.	1	2	4			
11	Os dirigentes e chefes encorajam as unidades e seus líderes a incorporar Big Data e Governança de Dados a seus processos decisórios e de negócios.		4	1	2		
12	Os dirigentes e líderes que estão na instituição há mais tempo utilizam o Big Data e a Governança de Dados para orientar tanto suas decisões estratégicas quanto táticas.	1	4	2			
13	Os gestores das áreas do Órgão utilizam o Big Data e o Analytics para orientar suas decisões.		4	3			
14	O processo de priorização e alocação de nossos ativos de Big Data (dados, pessoal, software e hardware) é orientado e examinado pelo núcleo de altos dirigentes.	1	3	1	2		
<b>METAS</b>							
15	Priorizamos nossas iniciativas de Big Data e Governança de Dados de acordo com as oportunidades que geram mais valor para nos destacarmos em termos de resultados institucionais.	1	3	2		1	





#### **4.2.2 Comentários da Avaliação de Prontidão para Big Data – Grupo Focal**

##### **1. Quais as percepções sobre as perguntas apresentadas no contexto de DADOS?**

Comentários:

E muito comum encontrar muitos dados nos órgãos, o problema não são os dados, o problema vai mais na linha de extrair valor dos dados, muito se fala que o dado é o novo petróleo. Na verdade, não é o dado mais o que extrai de valor e criar algum conhecimento a partir desses dados.

Não faz muito sentido ter uma volumetria muita alta e não saber o que extrair dos dados e produzir valor.

##### **2. Quais as percepções sobre as perguntas apresentadas no contexto de NEGÓCIO DO ÓRGÃO?**

Comentários:

Existe um foco em capacitação de prestadores de serviços terceirizados para os órgãos públicos. E um direcionamento de projetos de treinamento e investimento de capacitação na implementação de Big Data de Governança de Dados. Muitos órgãos optam pela contratação de profissionais do mercado.

Os servidores públicos e prestadores de serviços terceirizados de Big Data do Tribunal de Contas da União – TCU e Câmara dos Deputados, são incentivados a realizarem capacitações nas áreas.

##### **3. Quais as percepções sobre as perguntas apresentadas no contexto de LIDERANÇA?**

Comentários:

Os processos são frágeis, falta de priorização e descontinuação de iniciativas após a troca da gestão e liderança, sem pesar a longo prazo.

Iniciativas de Big Data realizadas pelas equipes técnicas e sem a participação da alta gestão e liderança.

#### **4. Quais as percepções sobre as perguntas apresentadas no contexto de METAS?**

Comentários:

Os Órgãos Públicos não tem uma cultura de difundir iniciativas integradas de Big Data.

Na prática, alguns órgãos já estão pensando em novos processos de inovação e no entendimento de Data Driven para extrair conhecimento e solucionar problemas.

As oportunidades e iniciativas de Big Data existem, mas como ser feito e como tem que ser feito, experimento e mais iniciativa das equipes técnicas do desenvolvimento do que metas efetivas das instituições, ainda precisa evoluir e promover um novo mindset.

#### **5. Quais as percepções sobre as perguntas apresentadas no contexto de TECNOLOGIA?**

Comentários:

Os Órgãos estão utilizando soluções de tecnologias híbridas para atender necessidades específicas. E ferramentas ideais para lidar com ativos de Big Data.

Falta expertise para utilizar as ferramentas para resolver o problema.

#### **6. Quais as percepções sobre as perguntas apresentadas no contexto de EQUIPE DE TI E ANALISTAS DE DADOS?**

Comentários:

Carência de profissionais qualificados ainda é uma das maiores barreiras para a implantação de Big Data na gestão de TI entre os Órgãos Públicos. Embora esses desafios apresentem impasses, está mais claro que o investimento de Big Data e Gestão de Dados proporcionara maior eficiência e benefícios no serviço Público Federal.

### 4.2.3 Entendendo o cenário atual

Desta forma, é possível perceber através da pesquisa acima (seção 4.2.3), busca melhorara compreensão das dificuldades de praticar governança de dados e big data, bem como as carências, processos de melhorias citadas nos tópicos subjacentes.

Pode-se dizer que o big data está redefinindo o panorama do gerenciamento de dados, de processos de extração, transformação e carregamento a novas tecnologias (como *Hadoop*) para limpar e organizar dados não estruturados em aplicativos de big data.

Embora o setor de negócios seja líder no desenvolvimento de aplicativos de big data, o setor público começou a obter *insights* para ajudar a apoiar a tomada de decisões em tempo real a partir de dados em movimento de rápido crescimento de várias fontes, incluindo a Web, sensores biológicos e industriais, vídeo, e-mail e comunicações sociais.

Perecebe-se que a prática de governança de dados aliado a big data em governos podem ajudá-los a servir seus cidadãos e superar os desafios nacionais (como aumento dos custos de saúde, criação de empregos, desastres naturais e terrorismo), mas também há algum ceticismo quanto à possibilidade de realmente melhorar as operações do governo, visto que os governos devem desenvolver novas capacidades e adotar novas tecnologias (como *Hadoop* e *NoSQL*) para transformá-lo em informação por meio da organização e análise de dados.

Aqui, pergunta-se, se os governos são capazes de implementar alguns dos aplicativos de big data associados ao setor de carencia, como os supra-citados acima. Primeiro, compartilha-se os setores em termos de objetivos, missões, processos de tomada de decisão, atores de decisão, estrutura organizacional e estratégias, depois volta-se para várias aplicações atuais em tecnologias avançadas.

Desta forma, salienta-se sobre a real necessidade de se aplicar Governança de dados aliadas a boas práticas de Big Data, não só visando um conjunto de dados

estruturados, validos, com riquezas de informações governamentais de interesse público que podem e devem ajudar a sanar vários dos problemas que o Brasil vem enfrentando. E válido dizer que:

- Empresas, governos e a comunidade de pesquisa podem derivar valor da enorme quantidade de dados digitais que coletam;
- Os governos dos principais países de TIC iniciaram projetos de aplicação de big data para aumentar a eficiência operacional, transparência, bem-estar dos cidadãos e envolvimento em assuntos públicos, crescimento econômico e segurança nacional;
- A análise de projetos de aplicativos de big data por governos oferece orientação aos países seguidores para suas próprias iniciativas de big data futuras.

Tais observações citadas acima, são interessantes justamente porque expressam direitos e deveres enquanto órgãos, governos, comunidade, entre outros. As riquezas de informações que são coletadas hoje através dos grandes centros de redes distribuídas, principalmente na internet, ajudam na compreensão de vários fatores socio-econômicos, que se forem bem tratados, extraídos, compreendidos, podem poupar vários recursos, além de viabilizar novos investimentos, direcionados justamente a carência de cada setor.

#### **4.2.5 Guideline para implantação de Governança de dados e Big Data em Instituições públicas federais**

EM ELABORAÇÃO

## 5. CONCLUSÕES

A análise acima, teve como objetivo principal, relacionar as diretrizes sobre boas práticas e aplicações de Big Data para o setor público, buscando o entendimento sobre os objetivos e maiores desafios e dificuldades relacionados com a adoção da Big Data no cenário abordado.

Objetivos específicos consideraram um escopo mais focado em aspectos relacionados com as necessidades e com os papéis dos órgãos públicos mediante a sociedade administrada, o que foi estudado por meio das questões elencadas a seguir.

- Quais são os impactos econômicos das soluções de Big Data no setor público?
- Quais as oportunidades e desafios da adoção de Big Data no setor público?
- Quais os impactos das práticas de Big Data no setor público?

Assim, se buscou fornecer as seguintes contribuições:

- Fornecer dados concisos que comprovem que a prática de Big Data pode interferir de maneira positiva na Administração Pública Federal, através das boas práticas, aplicação de diretrizes consistentes, apoio e capacitação para os profissionais envolvidos;
- Contribuir com o estímulo de incremento de eficiência e eficácia dos processos e na análise dos dados dos setores públicos;
- Dissertar sobre os processos de Big Data disseminando boas práticas (ambiente integrado, disposição e estruturação dos dados, aplicação dos 5v's, segurança) na administração pública, contribuindo com literatura técnica e científica para o corpo acadêmico e futuros leitores que possam se interessar pelo tema.

Estas contribuições estão indiretamente ligadas a perceber os motivadores e os benefícios obtidos com as práticas de Big Data nas organizações públicas, em suma, em conformidade com os objetivos da dissertação. A revisão de literatura informou que, como motivadores à utilização de Big Data por órgãos governamentais, está a capacidade de melhoria de decisões, incremento da participação dos cidadãos, melhor processamento de dados, economia de recursos públicos, questões de transparência etc. Cita como desafio a questão do alinhamento de projetos de Big Data com a missão institucional. As oportunidades estariam fortemente relacionadas com os motivadores.

Mediante o fato de que pouco da literatura estava relacionada com o caso brasileiro, realizou-se a pesquisa de campo para verificar o alinhamento destas questões, oportunidades, desafios e motivadores de projetos de Big Data, nos órgãos públicos do Governo Federal.

Conforme relato no Capítulo 4, verificou-se primeiro se os entrevistados teriam conhecimento que os permitisse responder com segurança as questões centrais da dissertação quanto ao que ocorre nos órgãos públicos que representavam. Uma vez verificado que em sua maioria os respondentes eram pessoas que sabiam não apenas tratar do Big Data, mas tratar do uso do Big Data em suas instituições, se coletou, por meio de outras questões, respostas para as questões centrais.

Quanto aos impactos econômicos do Big Data se verificou que existem impactos econômicos, em sua maioria positivos, com redução de gastos de recursos públicos, embora em parte considerável das respostas não se tenha verificado segurança quanto ao fato. Isso pode ter ocorrido por parte do público ainda relacionar de forma muito indireta o gasto de recursos do órgão com gastos de recursos em TI. No entanto, o caso brasileiro, ao menos com base nas respostas recebidas tendeu de forma pouco intensa a se alinhar à ideia constante dos casos da literatura revisada de que o uso do Big Data tende a promover a redução de recursos públicos.

As oportunidades e desafios na Adoção do Big Data estavam muito concentrados em aspectos da missão institucional, porém com parte dos desafios

impactados por questões políticas e administrativas dos órgãos. Os impactos, por sua vez, além dos econômicos estariam ligados a uma melhoria na tomada de decisão.

Em todos os momentos os respondentes deixaram claro que o aspecto da missão institucional não vem sendo esquecido pelos órgãos que possuem projetos e ambientes de Big Data, o que se alinha, mais uma vez com o caso da literatura. No entanto, pareceu haver pouco destaque na maioria das respostas aos aspectos de transparência e participação do cidadão, embora a literatura tenha sido em rica em citar estas questões quando trata de Big Data no setor público.

Um aspecto que chamou a atenção é que mediante a quantidade de pessoas que recebeu o formulário de pesquisa, a quantidade de respostas tenha sido baixa. Há a percepção, não comprovada, de que a crise sanitária decorrente do surto da COVID-19 possa ter impacto de alguma forma os profissionais a não atentarem para a contribuição da pesquisa.

Uma resposta possível ao que o objetivo principal da pesquisa buscava analisar é que, portanto, os desafios quanto à implantação de Big Data se relacionam com a melhoria de tomada de decisão mediante os objetivos institucionais de cada órgão público brasileiro representado, objetivando melhoria de eficiência no gasto de recursos públicos, não se olvidando das questões de transparência, embora sem o mesmo destaque se comparado às questões econômicas em si, mas tendo os projetos de Big Data enfrentado, ainda, muitas barreiras burocráticas e políticas.

No entanto, como se verificou, a maioria dos órgãos que possui ambientes e projetos de Big Data já observou que o alinhamento à missão institucional é o primeiro passo. Por sua vez se entende que este alinhamento é capaz de melhorar a superação do desafio causado pelas questões políticas.

Quanto à contribuição deste trabalho com os ambientes profissionais de Big Data no setor público, se verificou que, ao menos para novos projetos, para órgãos que ainda não possuam projetos de Big Data, ou que os possuam de maneira muito incipiente, pode-se verificar melhor o alinhamento com a missão institucional, mas também melhorar os estudos de impactos, inclusive os econômicos, uma vez que

estes estudos mais detalhados também podem auxiliar a derrubar as barreiras políticas. Os novos projetos podem também considerar melhor os aspectos de transparência e participação do cidadão, pois pouco se verificou o destaque de tais fatores nos projetos apurados. Estes aspectos também podem ser elencados de maneira a superar os desafios políticos ou ao menos reduzi-los. A disseminação desse conhecimento de estruturação de projetos com maior detalhamento de informações entre os prestadores de serviço de TI pode gerar uma melhoria do entendimento do escopo dos projetos não apenas do ponto de vista tecnológico, mas também do ponto de vista da contribuição social que os servidores de TI podem trazer para a comunidade administrada, melhorando inclusive seu comprometimento com cada projeto. Ou seja, trata-se de mais um impacto dos projetos de Big Data, muito positivo.

Como trabalho futuro pode-se ampliar o escopo da pesquisa a mais órgãos e colaboradores, bem como fazer a pesquisa de desafios e de oportunidades de forma mais detalhada. Ainda, como trabalho derivado, se poderia estudar melhor projetos que tratem mais de questões de transparência e de participação do cidadão. A própria compreensão do porque os profissionais de TI não terem dado tanto destaque nestes itens pode se constituir em um bom escopo de pesquisa.



## 6. REFERENCIAL BIBLIOGRÁFICO

ALGMIN, A. and J. ZAINO (2018), ***Trends in Data Governance and Data Stewardship***: A 2018 DATAVERSITY Report, DATAVERSITY Education, LLC, [http://content.dataversity.net/rs/656-WMW-918/images/Trends%20in%20Data%20Governance%20and%20Stewardship\\_FinalRP-Graphs.pdf](http://content.dataversity.net/rs/656-WMW-918/images/Trends%20in%20Data%20Governance%20and%20Stewardship_FinalRP-Graphs.pdf) Acessado em: 06 Ago. 2020.

BALTIMORE (2019). **Mapa dos desertos alimentares de Baltimore City**. Disponível em: [https://translate.googleusercontent.com/translate\\_c?depth=1&hl=pt-BR&prev=search&rurl=translate.google.com.br&sl=en&sp=nmt4&u=https://www.baltimoresun.com/maryland/baltimore-city/bal-bmorefoodmap-graphic-20150610-htmlstory.html&usq=ALkJrhiGA5F5zvCldW-OTPLTdbHxtgZhAQ](https://translate.googleusercontent.com/translate_c?depth=1&hl=pt-BR&prev=search&rurl=translate.google.com.br&sl=en&sp=nmt4&u=https://www.baltimoresun.com/maryland/baltimore-city/bal-bmorefoodmap-graphic-20150610-htmlstory.html&usq=ALkJrhiGA5F5zvCldW-OTPLTdbHxtgZhAQ). Acesso em: 12 mai. 2020.

BARATA, André Montoia. **Governança de dados em organizações brasileiras: uma avaliação comparativa entre os benefícios previstos na literatura e os obtidos pelas organizações**. 2015. Disponível em: <http://www.teses.usp.br/teses/disponiveis/100/100131/tde-28072015-215618/pt-br.php> . Acesso em: 06 Ago. 2020.

BARC (2019), ***Data Governance: Definition, Challenges & Best Practices***, Bi-Survey.com, <https://bi-survey.com/data-governance> Acessado em: 06 Ago. 2020.

BOCCATO, V. R. C. **Metodologia da pesquisa bibliográfica na área odontológica e o artigo científico como forma de comunicação**. Rev. Odontol. Univ. Cidade São Paulo, São Paulo, v. 18, n. 3, p. 265-274, 2006.

CARVALHO, Mónica Isabel Machado. **Data Governance: estudo e aplicação na EDP distribuição**. 2012. Faculdade de Economia, Universidade de Évora, Évora, 2012. 57 p. Disponível em: <http://hdl.handle.net/10316/21346> . Acesso em: 06 Ago. 2020

COMISSÃO EUROPEIA (1998). COM (1998). ***Public sector information: A key resource for Europe. Green paper on public sector information in the information society***. European Commission. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/01972240500388180> Acesso em: 09 mar. 2020.

CORDEIRO, Daniel. **Apache Hadoop: Conceitos técnicos e práticos, evolução e novas possibilidades**. 2012. Departamento de Ciências e Computação, Instituto de Matemática e Estatística – Universidade de São Paulo. Disponível em: <http://www.each.usp.br/dc/papers/erad-hadoop-DanielCordeiro.pdf> Acesso em: 01 mar. 2020.

CORREIA, Z. P. (2004). ***Toward a stakeholder model for the co-production of the public-sector information system***. *Information Research*, 10(3), paper 228. Retrieved February 27, 2013. Disponível em: <http://InformationR.net/ir/10-3/paper228.html>. Acesso em: 09 mar. 2020.

DANA-PMBOK (2018) ***Navigating the Labyrinth is a concise guide to data management, providing an executive's overview to the DMBOK2***. Utgitt 2018-06-01. Disponível em: [https://doc.difi.no/nasjonal-arkitektur/kunnskap\\_bok-navigating-the-labyrinth/](https://doc.difi.no/nasjonal-arkitektur/kunnskap_bok-navigating-the-labyrinth/) Acesso em: 12 ago. 2020.

DATA GOVERNANCE INSTITUTE (DGI). ***Definitions of Data Governance***. [S.l.]: DGI, c2017a. Disponível em: [http://www.datagovernance.com/adg\\_data\\_governance\\_definition/](http://www.datagovernance.com/adg_data_governance_definition/). Acesso em: 06 Ago. 2020.

DAVENPORT, Thomas H., 1954-***Big data at work: dispelling the myths, uncovering the opportunities***. pages cm: ISBN 978-1-4221-6816-5.

DELOITTE (2019). ***Using advanced analytics to drive regulatory reform***. Disponível em: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/public-sector/us-ps-using-advanced-analytics-to-drive-regulatory-reform.pdf> Acesso em: 12 mai. 2020.

DIAS, Jhenny Kelly do C. ***Big Data: Conceitos Básicos***. Disponível em: Acesso em: <http://www.cogef.ms.gov.br/wp-content/uploads/2018/04/Apresentac%C3%A7%C3%A3o-conceitual-do-Big-Data.pdf> 29 fev. 2020.

DRUID (2019). ***Apache Druid is a high performance real-time analytics database***. Disponível em: <https://druid.apache.org/technology> Acesso em: 10 mar. 2020.

ESPINDOLA, Priscilla Lüdtkke. ***Governança de dados aplicada à ciência da informação: análise de um sistema de dados científicos para a área da saúde***. Universidade do Estado de Santa Catarina. 2017. Disponível em: <file:///C:/Users/wesle/AppData/Local/Temp/8651080-Texto%20do%20artigo-42059-7-10-20180831.pdf> Acesso em: 06 Ago. 2020.

FAGUNDES, Priscila Basto., MACEDO, Dyllon Douglas J., FREUND, Gislaine Parra. ***A produção científica sobre qualidade de dados em Big Data: Um estudo na base de dados WEB of Science***. Universidade Federal de Santa Catarina, 2017. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/download/8650412/pdf/> Acesso em: 01 mar. 2020.

FILHO, Osilmar, M. C., CARDOSO, Marcelo de Castro. ***Arquitetura em Big Data para organização de milhões de ligações em uma empresa de Telecom***. Curso de

Engenharia de Computação Centro Universitário de Anápolis (UniEvangélica) – Anápolis, GO – Brasil. 2019. Disponível em: [http://repositorio.aee.edu.br/bitstream/aee/319/1/TCC2\\_2016\\_2\\_OsilmarMendonca.pdf](http://repositorio.aee.edu.br/bitstream/aee/319/1/TCC2_2016_2_OsilmarMendonca.pdf) Acesso em: 01 mar. 2020.

GAMA, João & RODRIGUES, Pedro. (2007). **Data Stream Processing**. 10.1007/3-540-73679-4\_3. Disponível em: [https://www.researchgate.net/publication/236007656\\_Data\\_Stream\\_Processing/citation/download](https://www.researchgate.net/publication/236007656_Data_Stream_Processing/citation/download) Acesso em: 10 mar. 2020.

GARTNER (2016) **GRC Redefined: Gartner Adopts an Integrated Risk Management View**. Disponível em: <https://blogs.gartner.com/john-wheeler/grc-redefined-gartner-adopts-an-integrated-risk-management-view/>. Acesso em: 29 julho 2020.

GHAVAMI, P. (2015), **Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL & Big Data Analytics**, CreateSpace Independent Publishing Platform.

GONÇALVES, Tânia Carolina Nunes Machado., VARELLA, Marcelo D. **Os desafios da Administração Pública na disponibilização de dados sensíveis**. Escola De Direito De São Paulo Da Fundação Getulio Vargas Revista Direito Gv | São Paulo | V. 14 N. 2 | 513-536 | Maio-Ago 2018. Disponível em: <http://www.scielo.br/pdf/rdgv/v14n2/1808-2432-rdgv-14-02-0513.pdf> Acesso em: 09 mar. 2020.

GONDINE, Sônia M. G. **Grupos focais como técnica de investigação qualitativa: desafios metodológicos**. 2002. Disponível em: <https://www.scielo.br/pdf/paideia/v12n24/04.pdf> Acesso em: 13 ag. 2020.

GUPTA, Anish., GUPTA, Manish K., **HIVE- Processing Structured Data in HADOOP**. International Journal of Scientific & Engineering Research Volume 8, Issue 6, June-2017 45 ISSN 2229-5518. Disponível em: [https://www.researchgate.net/publication/318338588\\_HIVE-Processing\\_Structured\\_Data\\_in\\_Hadoop](https://www.researchgate.net/publication/318338588_HIVE-Processing_Structured_Data_in_Hadoop) Acesso em: 10 mar. 2020.

HENRIQUES, Daniela A. COSTA, Helena R. **Big data – como utilizar a extraordinária quantidade de informações coletadas por novas tecnologias para obter vantagens competitivas**. UNA – 2018. Disponível em: [http://revistapensar.com.br/tecnologia/pasta\\_upload/artigos/a69.pdf](http://revistapensar.com.br/tecnologia/pasta_upload/artigos/a69.pdf). Acesso em: 06 Ago. 2020.

KAFKA, Apache. **O que é o Apache Kafka?** 2019. Disponível em: <http://kafka.apache.org/>. Acesso em: 18 mar. 2020.

KLEPPMANN, Martin. Making Sense of Stream Processing: The Philosophy Behind Apache Kafka and Scalable Stream Data Platforms. 2016. Disponível em: [https://assets.confluent.io/m/2a60fabedb2dfbb1/original/20190307-EB-Making\\_Sense\\_of\\_Stream\\_Processing\\_Confluent.pdf](https://assets.confluent.io/m/2a60fabedb2dfbb1/original/20190307-EB-Making_Sense_of_Stream_Processing_Confluent.pdf) Acesso em: 10 mar. 2020.

LADLEY J. **Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program**. The Morgan Kaufmann Series on Business Intelligence. Morgan Kaufmann. 2012.

MAGHAM, Ravi. **Pro Apache Phoenix: na SQL driver for Hbase**. 2017. Disponível em: [https://www.academia.edu/40508955/PRO\\_APACHE\\_PHOENIX\\_AN\\_SQL\\_DRIVER\\_FOR\\_HBASE](https://www.academia.edu/40508955/PRO_APACHE_PHOENIX_AN_SQL_DRIVER_FOR_HBASE) Acesso em: 18 mar. 2020.

MAY, T. **Pesquisa social: questões, métodos e processo**. Porto Alegre, Artmed, 2004.

MCKINSEY GLOBAL INSTITUTE. (2011, June). **Big Data: The next frontier for innovation, competition, and productivity**. McKinsey & Company. Disponível em: <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>. Acesso em: 09 mar. 2020.

NAOPA (2018). National Academy of Public Administration. **Strengthening Organizational Health and Performance in Government**. Disponível em: [https://www.napawash.org/uploads/Strengthening\\_Organizational\\_Health\\_and\\_Performance\\_in\\_Government.pdf](https://www.napawash.org/uploads/Strengthening_Organizational_Health_and_Performance_in_Government.pdf) Acesso em: 12 mai. 2020.

NARKHEDE, Neha., SHAPIRA, Gwen., PALINO, Todd. **Kafka: The Definitive Guide**. 2017. Disponível em: <https://book.huihoo.com/pdf/confluent-kafka-definitive-guide-complete.pdf> Acesso em: 18 mar. 2020.

NETFLIX (2019). Spark and Spark Streaming at Netflix. Disponível em: <https://www.slideshare.net/SparkSummit/spark-and-spark-streaming-at-netflix-sedakar-daxini> Acesso em: 18 mar. 2020,

OECD (2018), **Open Government Data Report: Enhancing Policy Maturity for Sustainable Impact**, OECD Digital Government Studies, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264305847-en>. Acesso em: 06 Ago. 2020.

OECD (2019), **Digital Government Review of Argentina: Accelerating the Digitalisation of the Public Sector**, OECD Publishing, Paris, <https://doi.org/10.1787/354732cc-en> . Acesso em: 06 Ago. 2020.

ORACLE (2019) **O que é Big Data?** Disponível em: <https://www.oracle.com/br/big-data/guide/what-is-big-data.html> Acesso em: 29 fev. 2020.

PETRY, Marcio. (2016). **Big Data e a implementação de um Sistema Distribuído com Apache Hadoop: um estudo exploratório**. Disponível em:

[https://www.researchgate.net/publication/318600685\\_Big\\_Data\\_e\\_a\\_implementacao\\_de\\_um\\_Sistema\\_Distribuido\\_com\\_Apache\\_Hadoop\\_um\\_estudo\\_exploratorio/citation/download](https://www.researchgate.net/publication/318600685_Big_Data_e_a_implementacao_de_um_Sistema_Distribuido_com_Apache_Hadoop_um_estudo_exploratorio/citation/download) Acesso em: 01 mar. 2020.

PHOENIX, Apache. **OLTP e análise operacional para Apache Hadoop**. 2019. Disponível em: <https://phoenix.apache.org/> Acesso em: 18 mar. 2020.

PREDPOL (n.d.). **Retrieved September 08, 2013, from PredPol** Web site: <https://www.predpol.com/>. Acesso em: 09 mar. 2020.

QBAL, Muhammad & SOOMRO, Tariq. (2015). **Big Data Analysis: Apache Storm Perspective**. International Journal of Computer Trends and Technology. 19. 9-14. 10.14445/22312803/IJCTT-V19P103. Disponível em: [https://www.researchgate.net/publication/271196175\\_Big\\_Data\\_Analysis\\_Apache\\_Storm\\_Perspective/citation/download](https://www.researchgate.net/publication/271196175_Big_Data_Analysis_Apache_Storm_Perspective/citation/download) Acesso em: 18 mar. 2020.

REIS, Rodrigo Xavier. **A utilização do fenômeno Big Data na Administração Pública A experiência do PENSA na Prefeitura Municipal do Rio de Janeiro**, 2015. Disponível em: <https://www.tce.rj.gov.br/documents/454798/49753202/Rodrio%20Xavier%20dos%20Reis.pdf> Acesso em: 09 mar. 2020.

ROSSOUW, L. **Big Data – grandes oportunidades**. Gen Re – Risk Insights, vol. 16, nº 2, 2012.

SANTANA, Lucas Vasconcelos. **Introduzindo o Apache Storm MAC5742 - Computação Paralela e Distribuída** Instituto de Matemática e Estatística Universidade de São Paulo. 2015. Disponível em: <https://www.ime.usp.br/~gold/cursos/2015/MAC5742/reports/ApacheStorm.pdf> Acesso em: 18 mar. 2020.

SANTOS, Isabel Maria Francisca dos. **Uma proposta de governança de dados baseada em um método de desenvolvimento de arquitetura empresarial**. 2010. 140 p. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2010. Disponível em: <http://www2.uniriotec.br/ppgi/banco-de-dissertacoes-ppgi-unirio/ano-2010/uma-proposta-de-governanca-de-dados-baseada-em-um-metodo-de-desenvolvimento-de-arquitetura-empresarial/view>. Acesso em: 06 Ago. 2020.

SENADO FEDERAL (2012). **Manual de Padronização de Atos Administrativos Normativos**. Disponível em: <https://www2.senado.leg.br/bdsf/bitstream/handle/id/496338/000960587.pdf?sequence=1&isAllowed=y> Acesso em: 09 mar. 2020;

SILVA, Bruno A. (2019). **Apache Hadoop**. Universidade Federal de São Carlos. Disponível em: <https://dcomp.sor.ufscar.br/verdi/topicosCloud/Hadoop.pdf> Acesso em: 01 mar. 2020.

SILVA, Petrônio. (2015). **Big Data e Data Science: Admirável Mundo Novo**. 10.13140/RG.2.1.2665.4806. Disponível em: [https://www.researchgate.net/publication/289253933\\_Big\\_Data\\_e\\_Data\\_Science\\_Admiravel\\_Mundo\\_Novo](https://www.researchgate.net/publication/289253933_Big_Data_e_Data_Science_Admiravel_Mundo_Novo) Acesso em: 29 fev. 2020.

SOUZA, Milton Pereira. **Mecanismos para viabilizar transparência ativa e abertura de dados adequados a governança de informações digitais da ANAC**. Universidade Federal de Minas Gerais. 2019. Disponível em: <https://repositorio.ufmg.br/bitstream/1843/30472/1/MiltonPereiradeSouza.pdf> Acesso em: 09 mar. 2020.

STORM, Apache. **Porque utilizar o Apache Storm**. 2019. Disponível em: <https://storm.apache.org/index.html> Acesso em: 18 mar. 2020.

STUMPF, Ricardo D. **O porquê de governança de dados em organizações de controle**. TCU – 2017. Disponível em: <https://webcache.googleusercontent.com/search?q=cache:vJld4o5ftqYJ:https://revista.tcu.gov.br/ojs/index.php/RTCU/article/view/1383/1529+&cd=1&hl=pt-BR&ct=clnk&gl=br&client=firefox-b-d> Acesso em: 06 Ago. 2020.

TALHAOUI, Mohamed Amine. (2018). **Real-time Data Stream Processing - Challenges and perspectives**. International Journal of Computer Science Issues. 14. 10.20943/01201705.612. Disponível em: [https://www.researchgate.net/publication/326985824\\_Real-time\\_Data\\_Stream\\_Processing\\_-\\_Challenges\\_and\\_Perspectives/citation/download](https://www.researchgate.net/publication/326985824_Real-time_Data_Stream_Processing_-_Challenges_and_Perspectives/citation/download) Acesso em: 10 mar. 2020.

TEHSEEN (2016). **Big Data e os 3 Vs**. Disponível em: <http://tehseen.dbsdataproyects.com/wp-content/uploads/sites/107/2016/04/3vs1.jpg>, Acesso em: 28 fev. 2020.

VAN KASTEREN, T., ULRICH, B., SRINIVASAN, V., & NIESSEN, M. (2014). **Analyzing tweets to aid situational awareness**. 36th European Conference on Information Retrieval. Disponível em: [https://link.springer.com/chapter/10.1007/978-3-319-06028-6\\_79](https://link.springer.com/chapter/10.1007/978-3-319-06028-6_79). Acesso em: 09 mar. 2020.

WANG, T., RUDIN, C., WAGNER, D., & SEVIERI, R. (2013). **Detecting patterns of crime with series finder**. Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Disponível em: <https://www.aaai.org/ocs/index.php/WS/AAAIW13/paper/viewPaper/7018> Acesso em: 09 mar. 2020.

YANG, Fangin., TSCHETTE, Eric., LÉAUTÉ, Xavier., **Druid: A real time analytical data store**. Disponível em: <http://static.druid.io/docs/druid.pdf> Acesso em: 10 mar. 2020.



YIU, C. (2012). ***The Big Data Opportunity. Making government faster, smarter and more personal.*** London: Policy Exchange. Disponível em: <http://ict-industry-reports.com.au/wp-content/uploads/sites/4/2013/05/2012-The-Big-Data-opportunity-in-Government-UK-Policy-Exchange-July-2012.pdf> Acesso em: 21 julho 2020.

ZILLNER, S., RUSITSCHKA, S., MUNNÉ, R., LIPPELL, H., LOBILLO, F., HUSSAIN, K. et al. (2013): ***D2.3.1. First draft of the sectorial requisites.*** Public Deliverable of the EU-Project BIG (318062; ICT-2011.4.4). Disponível em: [http://big-project.eu/sites/default/files/D2.4.1\\_FINAL\\_v0\\_8.pdf](http://big-project.eu/sites/default/files/D2.4.1_FINAL_v0_8.pdf) Acesso em: 09 mar. 2020.

ZIMMERMANN, Marlene Harger., MARTINS, Pura Lúcia Oliver. **Grupo focal na pesquisa qualitativa: relato de experiencia.** 2008. Disponível em: [https://educere.bruc.com.br/arquivo/pdf2008/211\\_86.pdf](https://educere.bruc.com.br/arquivo/pdf2008/211_86.pdf). Acesso em: 13 ago. 2020.

GILBERT, MJ. The antropologist as alcoholologist: qualitatitive perspectives and methods in alcohol research. *Int J Addict*, v.25, n.2A, p.127-48, 1990-91.

MOSCAROLA, J. *Enquêtes et analyse de données.* Paris: Vuibert, Gestion, 1991

## APÊNDICE A – Mapeamento Sistemático

### **Mapeamento Sistemático sobre Uma Análise do Impacto das Práticas de Big Data na Administração Pública Federal**

Fernando Gonçalves de Oliveira  
*Mestrando Profissional em Engenharia de Software*  
*Centro de Estudos e Sistemas Avançado do Recife - C.E.S.A.R*  
*Recife, PE - Brasil*  
*E-mail: [fernando.infinite@gmail.com](mailto:fernando.infinite@gmail.com)*

**Resumo** - Este artigo apresenta o mapeamento sistemático da literatura referente uma análise do impacto das práticas de Big Data na Administração Pública Federal, a qual busca obter subsídios para responder quais os objetivos e maiores desafios acerca da implantação de Big Data e/ou Ecossistemas de Big Data. O mapeamento sistemático está relacionado com o conhecimento da literatura correlata que pode auxiliar no entendimento da questão em epígrafe. O levantamento dos artigos foi realizado em três bases de dados (ACM Digital Library, IEEE Xplore e Science Direct - Elsevier) para as quais foi realizada uma série de filtros de maneira que se obtivesse um conjunto de artigos que forneçam as respostas para a questão de pesquisa elencada. Compreendeu-se que a implantação de Big Data no setor público precisa estar perfeitamente alinhada com a missão institucional do órgão que se propõe a implantar Big Data, mas que os gestores públicos nem sempre percebem suas instituições como aptas a implantarem projetos de Big Data, o que pode, por sua vez, demandar que sejam apresentadas todo um conjunto de diferentes projetos de Big Data em diferentes esferas e missões institucionais. Finalmente se percebe que há pouca literatura relacionada com o caso brasileiro, o que se pode destacar a importância, tanto no mapeamento sistemático quanto na pesquisa.

**Palavras-Chave:** Big Data, Administração Pública Federal, Ecossistema, Mapeamento Sistemático.

#### **1. Introdução**

Atualmente, indivíduos e organizações, sejam privadas, públicas ou do terceiro setor, vêm gerando grandes volumes de dados, em uma proporção muito grande e crescente AKOKA et al., [1]. Complementarmente, a popularização dos

dispositivos conectados (Internet of Things – IoT), sensores e aplicações que coletam e geram dados em forma de imagem, áudio e vídeo, têm contribuído significativamente para esse volume e sua curva ascendente. Esse contexto ganha impulso, relevância e complexidade com o advento da quarta revolução industrial, conhecida como Indústria 4.0 AKOKA et al., [1]; ALMEIDA et al., [2].

Big Data é um termo genérico para a enorme quantidade de dados digitais não estruturados, normalmente coletados de múltiplas fontes; o termo também representa os novos desafios que envolvem a complexidade, a segurança e os riscos à privacidade, bem como a necessidade de novas tecnologias e habilidades humanas, redefinindo o cenário do gerenciamento de dados e dos processos de extração, transformação e carregamento KIM et al., [3].

Akoka et al. [1] realizaram um trabalho de enfoque meta analítico com artigos sobre Big Data, o qual demonstra que a quantidade de estudos sobre o termo vem crescendo, o que pode ser relacionado com o aumento de percepção acerca da importância de tal conjunto de tecnologias e de sua implantação. Para responder à questão sobre quais os assuntos produzidos a partir das pesquisas de Big Data, os autores classificaram os trabalhos pesquisados em domínios de conhecimento, abrangendo, entre outros, energia, medicina, ecologia, finanças e governo. A presença do domínio “governo” demonstra a relevância do tema em relação a pesquisas com Big Data. Entretanto, seu posicionamento em oitavo domínio com mais trabalhos publicados, em um universo de doze domínios (com apenas 35 trabalhos pesquisados, de um total de 441 trabalhos analisados), evidencia a avenida de oportunidades e a carência de mais trabalhos relacionados ao domínio “governo” ou “setor público”.



No caso brasileiro, encontrou evolução o entendimento de que os órgãos do governo federal brasileiro devem participar de um ambiente que permita a adequada transparência das atividades públicas, bem como a necessidade de controle e conhecimento de informações em larga escala para melhor gestão das atividades relacionadas com os serviços públicos e governamentais, conforme é possível verificar no trabalho de Victorino et al., [4]. Os autores complementam seu trabalho apresentando uma proposta de Ecossistema de Big Data para dar suporte à análise de dados abertos governamentais como resposta ao contexto de transparência.

Este artigo tem como tema o mapeamento sistemático referente à análise do impacto das práticas de Big Data na Administração Pública Federal do Brasil, abordando o tema por meio de pensamentos de autores que delineiam as bases para o estabelecimento do ecossistema de Big Data e as boas práticas relacionadas. Assim, o objetivo principal é entender quais os propósitos e os maiores desafios e dificuldades relacionados com a adoção do Big Data na Administração Pública Federal (APF). A pesquisa permite identificar as iniciativas e a visão dos ecossistemas de Big Data que possa auxiliar ambientes profissionais ou organizacionais com aplicabilidade prática. O estudo em tela atinge, inicialmente, esse objetivo, sendo o ambiente profissional, no caso, a APF.

O restante deste artigo está estruturado da seguinte forma: a Seção 2 aborda alguns conceitos centrais para o entendimento da terminologia que compõe o cenário de Big Data e a Seção 3 apresenta a Metodologia da pesquisa. Na Seção 4 são apresentadas as conclusões deste mapeamento sistemático.

## 2. Conceitos

Nessa seção se revisa os principais conceitos referentes a Big Data e ao Ecossistema de Big Data, com o objetivo de fornecer um entendimento geral do cenário de estudo.

Segundo o artigo de Victorino et al., [4], citando Schimdt [5], a humanidade criou, até o ano de 2003, cerca de cinco exabytes de dados. O volume, desde então, dobra a cada dois anos em média VICTORINO et al., [4]; GANTZ; REINSEL, [6], o que fez com que no ano de 2019 fossem criados cerca de 160 exabytes de dados. Consolidando o conceito de Big Data, nas palavras de Araújo Júnior

e Souza [7], fundamentados em Ramos [8], Big Data é um conjunto de tecnologias utilizadas para tratar grandes volumes de dados. É importante destacar que, não necessariamente, os dados tomados para tratamento estão em formatos similares, compatíveis. Em um cenário como este, conforme destacam os mesmos autores, tornam-se necessárias soluções computacionais mais efetivas em termos de custos e mais inovadoras no tratamento da informação.

Victorino et al., [4] destacam como uma destas soluções computacionais o Big Data, o qual possibilita a gestão de dados em grande velocidade de processamento, mas que não se refere especificamente à sua organização. Araújo Júnior e Souza [7] destacam que o conceito de Big Data tem sua origem no Business Intelligence (BI), que é, por sua vez, resultado do processo de mineração de dados analíticos, voltado para a combinação de dados, formatação de metadados e de conteúdo a serem usados no processo decisório. O BI é o uso de variadas fontes de dados e informações utilizadas para definir estratégias e subsidiar processos organizacionais. Logo, os conceitos de Big Data e BI convergem, e por isso alguns autores consideram as duas concepções como apenas uma ARAÚJO JÚNIOR E SOUZA, [7]; BARBIERI, [9].

Araújo Júnior e Souza [7] explicam, ainda, alguns outros autores, tais como Davenport [10], Márquez and Marques e Lev [11] e Immon and Linstedt [12], que também consideram os conceitos similares. Entretanto, não se verifica que os autores se preocupem em citar a organização dos dados utilizados nos processos, nem mesmo acerca do armazenamento. No entanto, conforme Immon and Linstedt [12], os conceitos são similares no que tange a modelos que envolvem captura, apresentação e geração de relatórios para a tomada de decisão.

Davenport [10], citado por Victorino et al., [4], afirma que o termo Big Data é muito genérico, utilizado para conjuntos de dados que não são comportados em repositórios habituais. Logo, o termo estaria relacionado a conjuntos de dados massivos e volumosos. Não caberiam, portanto, em servidores tradicionais e bancos de dados relacionais, demandando novas tecnologias de armazenagem e tratamento.

É comum nos artigos analisados que os autores abordem a definição de '3Vs' AKOKA ET AL., [1]; ARAÚJO JÚNIOR E SOUZA, [7]; DAVENPORT, [10]; LANEY, [13], caracterizando o Big Data, quais

sejam, conforme definições presentes em Araújo Júnior e Souza [7]:

- **Volume:** permite a modelagem e análise de extensos volumes de dados e o que será extraído e armazenado;
- **Velocidade:** fornece resultados de pesquisa com agilidade, em tempo compatível com a demanda, na qual os dados serão gerados, a rapidez na ação tomada sob os dados ou a taxa de mudança nos dados; e
- **Variedade:** processa e fornece conjuntos de dados em diferentes tipos e formatos.

A estes '3vs', Davenport [10] acrescentou, posteriormente, como destacam Victorino et al., [4], outros dois 'Vs', totalizando, então, os '5Vs' da definição do Big Data. São os outros dois 'Vs', também na definição de Araújo Júnior and Souza ARAÚJO JÚNIOR E SOUZA, [7]: HASHEM et al., [14]:

- **Veracidade:** processa e fornece conjuntos de dados fiéis aos dados processados e armazenados nos bancos, incertezas e confiabilidade dos dados; e
- **Valor:** agrega valor informacional; informação de valor claro para as organizações que as utilizam.

A compilação deste '5Vs' é graficamente representada na Figura 1.

Volume	Variedade	Velocidade	Veracidade	Valor
				
Quantidade de dados que são gerados a cada segundo	Tipos de dados que podem ser utilizados	Velocidade a qual os dados são gerados e transmitidos	Grau de organização e confiabilidade dos dados	Necessidade de se transformar dados em valor
Terabytes Registros Transações Arquivos	Estruturado Semiestruturado Não estruturado	Tempo real Streams Batch	Fidedignidade Autenticidade Origem Relevância Disponibilidade	Estatística Eventos Correlações Hipóteses

Figura1. Demonstração de Big Data por meio de 5Vs.

O Hadoop é um framework de computação distribuída, implementado em Java, voltado para clusters e processamento de grandes massas de dados, que após mapear os dados distribuídos e aplicar as transformações ou regras de negócio (processo MapReduce), consolida o resultado e apresenta a resposta ao usuário final ALMEIDA et al., [2]; O'DRISCOLL et al., [15]; KAZEMI, [16]. De acordo com JADHAV et al., [17], a estrutura do Hadoop consiste no sistema de arquivos distribuídos

do Hadoop (Hadoop Distributed File System – HDFS), e outros projetos como Mapa Reduce, Yarn, Hive, Ping, Mahout, Avro, Sqoop, Oozie, Chukwa, Flume e Zookeeper chamados de ecossistema de componentes que ajuda resolver problemas de processamento de Big Data, observada na Figura 2.

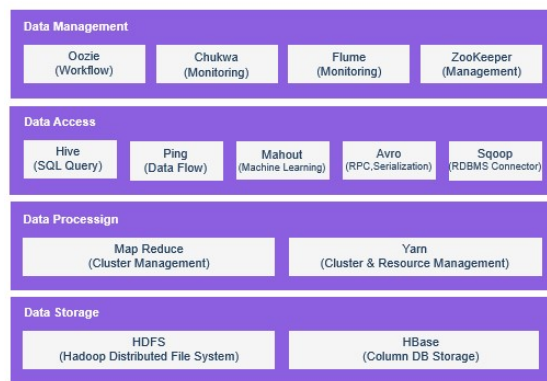


Figura 1. JADHAV [17] Hadoop Ecosytem Components, modelo adaptado pelo autor

Tal processamento de grandes massas de dados é um dos principais motivos pelos quais gestores têm buscado projetos de Big Data. Portanto, o provimento de capacidade analítica a partir da extração de novas visões derivadas de grandes volumes de dados, o que em bases de dados tradicionais pode ser considerado complicado, bem como as demais estruturas componentes compõem o Ecossistema de Big Data, conceito a seguir apresentado.

## 2.1. Ecossistema de Big Data

Para esta subseção retoma-se o exposto por Victorino et al., [4], o qual revisita os conceitos elencados em Shin e Choi [18], indicando que, assim como conceitos de ecossistemas em outras ciências, o conceito de ecossistema de Big Data estaria relacionado com as interações entre os diversos componentes de um sistema. No caso, estaria relacionado com as interações entre tecnologia, governo, indústria, mercados, usuários e sociedade, sendo examinados ou investigados os efeitos do ambiente de Big Data em todos os setores envolvidos. Victorino et al., [4] citam, ainda, a similaridade desta definição com a encontrada em Demchenko et al. [19], na qual a abordagem biológica é mantida, mas com a diferença na definição do ecossistema como um complexo de facilidades técnicas e componentes construídos à cerca de certa origem de dados específica e suas

aplicações, sendo que o complexo de componentes inter-relacionados é voltado para as atividades de armazenamento, processamento, visualização e entrega dos resultados. O ecossistema compreenderia, ainda, alguns componentes arquiteturais, tecnologias e outros componentes para o processamento e a análise do Big Data, devendo compreender o gerenciamento do ciclo de vida do Big Data, ou seja, o fluxo de transformação dos dados e ainda toda a infraestrutura de segurança do Big Data, bem como todo o conjunto necessário de componentes e políticas para prover controle no acesso aos dados ALMEIDA et al., [2].

É possível observar, se tomadas as explicações em epígrafe, de que os dados não se encontram em repositórios tradicionais, e que são oriundos de diversas fontes e, ainda, que o volume de dados é massivo ALMEIDA et al., [2], que Big Data e Cloud Computing (ou Computação em Nuvem, em língua portuguesa) estão relacionados. Esta relação é parte do trabalho de Hashem et al. [14], no qual os autores detalham que enquanto o Big Data fornece aos usuários a capacidade de utilizar poder computacional em larga escala para processar consultas distribuídas em vários conjuntos de dados e devolver as resultantes em tempo hábil, a Computação em Nuvem fornece o mecanismo subjacente por meio do uso de Hadoop.

O MapReduce DEAN; GHEMAWAT, [20] é um exemplo de processamento de Big Data em um ambiente de Computação em Nuvem, o qual permite o processamento de grandes quantidades de conjuntos de dados armazenados em paralelo no cluster Hadoop ALMEIDA et al., [2]; KAZEMI, [16]. A Computação em cluster apresenta bom desempenho em ambiente de sistema distribuído, com características tais como: baixo consumo de energia, bom aproveitamento de armazenamento e boa taxa de utilização e adaptação do uso da rede de computadores e de comunicações. Bollier et al. [21] destacam a capacidade da Computação em cluster de fornecer um contexto favorável ao crescimento de dados. Já Miller et al. [22] argumentam que a falta de disponibilidade de dados é cara porque os usuários transferem mais decisões para métodos analíticos; uso incorreto dos métodos ou pontos fracos inerentes nos métodos podem produzir decisões erradas com alto custo.

Estas intrincadas relações entre os componentes do Ecossistema de Big Data, incluindo os processos e ferramentas para o trabalho com as massas de dados, são destacadas por Almeida et al. [2]. No referido trabalho, se verifica a relação com a Computação em Nuvem/distribuída, bem como com

ferramentas e métodos de processamento de dados não utilizados na computação tradicional. Mais uma vez, até os sistemas de guarda de dados e a relação destes volumes com a era da informação e da demanda por alto poder computacional para decisões negociais são compreensíveis, além de ressaltar a complexidade do Ecossistema de Big Data.

Estas decisões e os processos para se obtê-las por meio da análise de grandes volumes de dados são, em uma análise, a razão da existência do Big Data. Razão esta que é reforçada no contexto da Administração Pública Federal, constituindo-se em ferramenta fundamental para apoiar os processos de tomada de decisões e de cumprimento dos normativos de transparência governamental.

## 2.2. Big Data no Setor Público

Conforme citado na Seção 1, o setor público carece da utilização de ferramentas de Big Data, seja por questões de melhor controle das informações para tomada de decisões, seja pela necessidade de transparência das ações estatais junto à coletividade administrada.

Mesmo com estas características, o setor público encontra diferenças com relação a implantação do Big Data e seus ecossistemas. Segundo Klievink et al. [23], o Big Data está sendo implementado com sucesso no setor privado e em pesquisas científicas. No entanto, o setor público parece estar ficando para trás. As organizações governamentais muitas vezes reconhecem as oportunidades relacionadas com Big Data, mas não se avaliam como aptas para a utilização de grandes volumes de dados, nem se consideram adequadamente equipadas para o Big Data. Os autores tratam exatamente acerca destas inseguranças e incertezas, apresentando uma estrutura para avaliar a prontidão das organizações públicas. O método utilizado busca desmistificar o conceito de Big Data para organizações do setor público holandês, conforme expresso em termos de características organizacionais específicas e mensuráveis. Os resultados obtidos sugerem que as organizações podem ser tecnicamente capazes de utilizar as tecnologias de Big Data, mas não obterão ganhos significativos com essas atividades se os aplicativos não se ajustarem às suas organizações e às principais tarefas estatutárias.

De acordo com Janssen e Van Den Hoven [24], considerando que além de serem massivos, os dados podem ser abertos, compondo o que é literariamente chamado de BOLD (do inglês, Big and Open Linked Data), a utilização de Big Data por entes

governamentais resulta em novas oportunidades e tem o potencial de transformar o governo e suas interações com o público. Como efeito colateral, o BOLD oferece a oportunidade de analisar até mesmo o comportamento dos indivíduos, aumentar o controle e reduzir a privacidade. Ao mesmo tempo, em linha com as melhores práticas, o BOLD pode ser usado para criar um governo transparente, reforçando a abordagem dos autores mencionados anteriormente, e com os objetivos de implantação do Big Data em entes governamentais. Ainda segundo Janssen e Van Den Hoven [24], transparência e privacidade são consideradas importantes valores sociais e democráticos que estão mudando a forma como os cidadãos informados participam de processos democráticos. Os conceitos de transparência e privacidade têm muitas interpretações e são considerados difíceis de serem compreendidos, o que dificulta, por consequência as formas de implementação dos sistemas relacionados com as informações detidas por entes estatais.

Considerando as premissas dos parágrafos anteriores, a motivação deste artigo é o estudo da implementação do Big Data no setor público brasileiro, buscando por trabalhos relacionados com os entes governamentais brasileiros. Passa-se, agora, à Seção 3 para entendimento dos procedimentos adotados.

### 3. Metodologia

#### 3.1. Introdução

Para atingir o objetivo proposto, o processo do Mapeamento Sistemático foi conduzido, conforme disposto na Figura 3.



Figura 3. Condução do mapeamento sistemático. Autoria própria.

Considerando os objetivos e os desafios da pesquisa citados anteriormente, observa-se como foco responder à questão principal (QP) a seguir.

QP) Quais os objetivos e maiores desafios ou dificuldades de implantação de Big Data no setor público?

Complementarmente, como perguntas secundárias (PS), que na sua integração alinhavam a resposta à questão principal, listam-se:

- PS1) Quais os impactos econômicos nas soluções de Big Data no setor público?
- PS2) Quais as oportunidades e desafios da adoção de Big Data no setor público?
- PS3) Quais os impactos que práticas de Big Data trazem ao desenvolvimento de soluções?

#### 3.2 Estratégia de pesquisa

Trata-se sobre mapeamento sistemático da literatura pautada na:

- 1) Elaboração de “strings” de pesquisa, orientadas pela estratégia de busca;
- 2) Variedade de fontes de pesquisa para localização dos estudos;
- 3) Definição de critérios de inclusão e exclusão de produções;
- 4) Avaliação da qualidade metodológica das produções recuperadas.
- 5) Bases de dados consultadas e estratégias de busca.

#### 3.3 Fontes de Pesquisa

O levantamento dos artigos foi realizado em três fontes de pesquisa, conforme exposto na Tabela 1.

Tabela 3. Bases selecionadas. Elaboração própria.

DATABASE	ENDEREÇO
ACM Digital Library	<a href="https://dl.acm.org/">https://dl.acm.org/</a>
IEEE Xplore	IEEE Xplore
ScienceDirect – Elsevie	<a href="https://www.sciencedirect.com/">https://www.sciencedirect.com/</a>

#### 3.4 Strings de Pesquisa

Considerando tais bases de pesquisa, algumas combinações de termos foram fundamentais para obtenção de artigos capazes de auxiliar no mapeamento sistemático e na obtenção do estado da

arte. Procedeu-se ao cruzamento das principais palavras-chave relacionadas aos temas investigados, quais sejam: “Big Data”, “Setor Público” e “Governo”, além de outros eventualmente necessários para o enriquecimento das fontes de pesquisa, com objetivo de abarcar uma maior quantidade de produções, evitando que algum estudo muito importante ou de muito destaque fosse excluído no levantamento. Assim, algumas “strings” de pesquisa foram montadas e todas as publicações selecionadas foram referentes ao período de 2016 a 2019, conforme apresentadas na Tabela 2.

Tabela 4. Strings de pesquisa. Elaboração própria.

ID	Base de dados	Query utilizada
1	ACM Digital Library	“query”: AllField: (“public sector”) AND AllField: (“Big Data”) AND AllField: (“difficulties” OR “opportunities” OR “economic impacts” OR “economic advantages”) “filter”: Publication Date: (01/01/2016 TO 12/31/2019)
2	IEEE Explore	(“Big Data” AND “Government” OR “Big Data” AND “Public Sector”)
3	ScienceDirect - Elsevier	“Big Data” AND “Government” AND “Public Sector”

### 3.5 Critérios de Seleção (Inclusão/Exclusão) de Artigos

Foram considerados alguns critérios de seleção para que um determinado artigo fosse incluído como passível de análise para a pesquisa, conforme apresentado na Tabela 3.

Tabela 5. Critérios de inclusão de artigos. Elaboração própria.

CI	Critério de inclusão de artigos
1	Apresenta os objetivos e maiores desafios ou dificuldades de implantação de Big Data
2	Aborda os impactos econômicos nas soluções de Big Data no Setor Público
3	Aborda as oportunidades e desafios da adoção de Big Data

De maneira semelhante, tomou-se alguns critérios para exclusão de artigos que apareceram entre os resultados das queries apresentadas na Subseção 3.4, quais sejam, os apresentados na Tabela 4.

Tabela 6. Critérios de exclusão de artigos. Elaboração própria.

CE	Critério de exclusão de artigos
1	Repetidos
2	Muito semelhantes
3	Não acessível
4	Não escrito em língua inglesa ou portuguesa
5	Publicado como “short paper” ou somente “pôster”
6	Sem resumo ou <i>abstract</i>
7	Publicado apenas resumo
8	Versão mais antiga de outro artigo já considerado ou incluído
9	Não auxilia na resposta de nenhuma das questões da pesquisa

### 3.6 Estudos dos Resultados

O procedimento descrito nas Subseções anteriores resultou numa quantidade de artigo por ano e por base de dados conforme apresentado na Tabela 5. As Tabelas 6 e 7 mostram a filtragem de artigos excluídos e selecionados, respectivamente, por base/ano.

Tabela 5. Artigos por base/ano. Elaboração própria.

Base de dados	2016	2017	2018	2019	2020	Total
ACM Digital Library	035	041	055	053	017	0201
IEEE Explore	193	233	254	257	082	1019
Science Direct - Elsevier	108	138	154	242	158	0800
Total Geral	336	412	463	552	257	2020

Tabela 6. Artigos excluídos por base/ano. Elaboração própria.

Base de dados	2016	2017	2018	2019	2020	Total
ACM Digital Library	035	041	055	053	017	0201
IEEE Explore	193	233	254	257	082	1019
Science Direct - Elsevier	108	138	154	242	158	0800
Total Geral	336	412	463	552	257	2020

Tabela 7. Artigos selecionados por base/ano. Elaboração própria.

Base de dados	2016	2017	2018	2019	2020	Total
ACM Digital Library	035	041	055	053	017	0201
IEEE Explore	193	233	254	257	082	1019
Science Direct - Elsevier	108	138	154	242	158	0800
Total Geral	336	412	463	552	257	2020

Para a priorização dos artigos, bem como das respostas das perguntas, foi aplicada a metodologia de enfoque meta-analítico, conforme os trabalhos de Mariano and Santos [25], Mariano et al. [26] e Sánchez-Meca and López-Pina [27], o qual tende a apresentar os trabalhos de maior relevância acadêmica ou com maior quantidade de citações acadêmicas em destaque dentro de um conjunto previamente obtido. Graficamente, as atividades de tal metodologia estão esquematizadas na Figura 4. Os detalhes de cada uma das três etapas representadas na aludida Figura 4 estão descritas em Mariano and Santos [25].

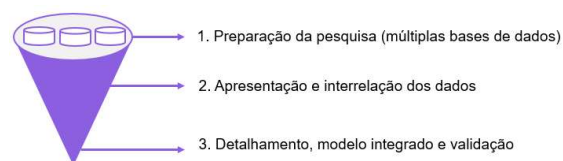


Figura 4. Diagrama Enfoque Meta Analítico. Extraída de Mariano and Santos [25].

No entanto, destaca-se que alguns artigos, não resultantes da priorização com o enfoque meta-analítico, bem como do procedimento principal descrito ao longo do presente trabalho, foram mantidos ou incluídos como base literária da pesquisa. Isso se deve ao fato de haver relação destes trabalhos com o caso brasileiro, como por exemplo o artigo de Victorino et al., [4], que trata especificamente de Big Data no setor público brasileiro, embora ainda encontre poucas citações ou relevância dentro das bases pesquisadas.

Com fundamento nos estudos selecionados, algumas contribuições podem ser identificadas, possibilitando inferir respostas às perguntas secundárias PS1, PS2 e PS3, relatadas na Subseção 3.1. Tal resultado pode ser obtido por meio de enfoque em certos termos componentes da busca principal. Os quantitativos de estudos selecionados, que apoiam as respostas às perguntas secundárias, estão sintetizados na Tabela 8.

Tabela 8. Bases selecionadas. Elaboração própria

Base de dados	PS1	PS2
ACM Digital Library	7	4
IEEE Explore	8	10
Science Direct - Elsevier	4	5

Por exemplo, para selecionar artigos capazes de auxiliar a resposta da PS1 (impactos econômicos), pode-se focar nos artigos que citam impactos econômicos e criação de valor para organizações, por meio da adoção de Big Data, no caso, no setor público. Com este diapasão, estão sendo identificadas contribuições como as de Khtira et al., [28] ou mesmo de Klievink et al. [23], as quais mencionam as possibilidades de criação de valor em diferentes áreas de atuação dos órgãos governamentais, em seus campos de atuação.

Igualmente, citam uma melhor eficiência quanto à análise e coleta de informações e na interação com a comunidade

administrada, permitindo, em última análise, melhor emprego dos recursos públicos ou mesmo sua economia.

Para a PS2 (oportunidades e desafios), foram encontradas contribuições no artigo de Klievink et al. [23], o qual relata entre as dificuldades de implementação do Big Data no setor público, inicialmente, a dúvida sobre a capacidade, incluindo aspectos de infraestrutura computacional e da adoção de ferramentas de Big Data. Em adição, é informado no aludido trabalho que possuir a infraestrutura e as ferramentas pouco têm a acrescentar se o trabalho de Big Data não estiver muito correlacionado com a missão institucional de um determinado órgão público ou governamental.

Ou seja, não acrescenta valor se não houver um projeto de Big Data conectado à missão institucional. Khtira et al. [28] citam ainda que o Big Data no setor público se torna uma necessidade também em virtude das novas maneiras de a sociedade interagir com os órgãos governamentais, o que já é capaz de gerar enormes quantidades de dados, as quais as organizações públicas devem saber manusear e tratar de forma a obter informação capaz de conduzir políticas públicas e orientar a tomada de decisão. As oportunidades de aplicações do Big Data para o setor público vão desde grandes aplicações em bases de dados de governos nacionais, passando pela aplicação de serviços públicos em comunidades locais, como exemplificado pelo trabalho de Malomo and Sena [29].

Finalmente, considerando a PS3 (impactos), as dificuldades na implementação do Big Data estão muito ligadas aos desafios, citados na PS2, em especial quando se consideram as organizações públicas. Inicialmente, o já citado artigo de Klievink et al., [23] aponta a questão da compreensão da capacidade de implementação, considerando aspectos de infraestrutura computacional. Igualmente, a dificuldade de construir um projeto de Big Data realmente capaz de auxiliar no acréscimo de valor adequado à missão institucional do órgão em questão. O artigo de Malomo e Sena [29] também enfatiza que as organizações públicas, neste caso representadas por seu corpo técnico-funcional, muitas vezes não dominam as questões sobre como o Big Data pode ser útil no crescimento de eficiência das atividades públicas. Os artigos não se furtam, porém, a citar também dificuldades técnicas, relacionadas, por exemplo, com a multiplicidade de fontes de dados, muitas vezes derivadas de dados não estruturados, principalmente em virtude da forma como são coletados.

A pesquisa, conforme mencionado em epígrafe, encontra considerável relevância, uma vez que as dificuldades de implementação do Big Data podem ser comuns a diversas entidades públicas e até mesmo privadas. As eventuais soluções, por sua vez, em virtude de muitas dificuldades serem comuns, podem se constituir em soluções que auxiliem, portanto, diversas organizações.

Tem-se, portanto, que a principal questão a ser tratada no trabalho ora proposto é: Quais os objetivos e maiores desafios e dificuldades de implantação do Big Data no setor público? Segundo Victorino et al., [4], os objetivos das organizações governamentais ao participar de um ambiente de Big Data estão diretamente relacionados com as suas funções de fornecimento de serviços públicos e com a necessidade de provimento de informações aos cidadãos, incluindo questões relacionadas com a transparência de governo, o que corrobora, mesmo que este autor tenha escrito seus trabalhos com vistas ao caso brasileiro, o explicitado pelos citados artigos utilizados na análise das questões descritas.

Graficamente, considerando os procedimentos descritos entre as Subseções 3.1 a 3.5, sintetiza-se os resultados conforme apresentados na Figura 5.

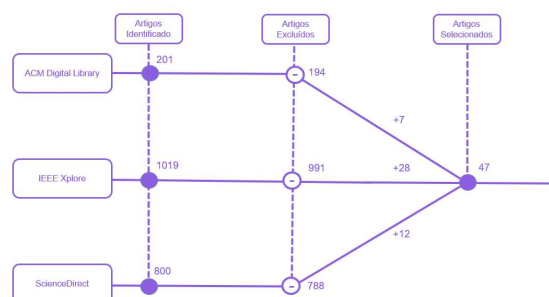


Figura 5. Síntese dos resultados dos procedimentos da pesquisa bibliográfica. Elaboração própria

#### 4. Discussões

Na revisão foram encontrados desafios para a implantação de Big Data no setor público AL-SAI; ABUALIGAH, [30]. Como forma de responder à questão principal deste trabalho, os objetivos e desafios estão classificados em perspectivas, conforme Tabela 9.

*Tabela 9. Desafios e possíveis soluções propostas.  
Adaptada de Al-Sai e Abualigah [30]*

Desafios	Soluções Possíveis
Perspectiva de Tecnologia	<p>Aumento da capacidade de armazenamento e utilização do armazenamento disponível, como armazenamento em nuvem para armazenamento e processamento de big data. Implementando questões de segurança questões de política, legislação e regulamentação.</p> <p>Usando as ferramentas eficazes de gerenciamento de dados para analisar dados de estrutura, desestruturação e estrutura simulada de várias fontes.</p> <p>Fornecendo e integrando recursos eficientes de big data.</p>
Perspectiva de Pessoas	<p>Aprimorando as habilidades do capital humano disponível em big data, como análise de dados, segurança de big data e importar o especialista que pode apoiar a implementação de big data.</p> <p>Implementando uma iniciativa nacional que apoia a colaboração entre big data e governo eletrônico.</p> <p>Aumentar a conscientização sobre big data e o impacto no governo eletrônico.</p> <p>Permitir que os cidadãos sejam mais criativos e se expressem efetivamente nas redes sociais.</p>
Perspectiva de Processos de Negócio	<p>Desenhe uma estratégia eficaz que suporte o roteiro para a implementação de big data no governo eletrônico.</p> <p>Estabelecer parcerias públicas e privadas.</p> <p>Crie uma rede entre comunidade e governo sob a égide de big data.</p> <p>Formule estratégias eficazes e um roteiro que leve a um ambiente de big data.</p> <p>A liderança deve motivar, envolver, influenciar o suporte à implementação de big data no governo eletrônico.</p> <p>Integre um grande volume de dados valioso no governo eletrônico aos processos de tomada de decisão.</p> <p>Capacitar o papel da e-Liderança e Gerenciamento da informação na formulação do plano estratégico e decisões.</p>

Além desses desafios existem outros obstáculos para o tema virar realidade dentro do setor público. Estudo de outros países, como no caso da Turquia, demonstram que o custo, infraestrutura insuficiente, questões de privacidade e segurança, burocracia governamental, falta de legislação sobre o tema, adaptação dos servidores ao processo e a integração dos sistemas de T.I existentes são grandes obstáculos para a efetiva adoção de soluções de Big Data DOGDU et al., [31].

#### 4. Conclusão

Este artigo apresenta um mapeamento sistemático sobre o tema de implantação do Big Data na Administração Pública do Brasil, para fundamentação de pesquisa acadêmica que busca verificar como os órgãos do Governo Brasileiro têm implantando as aludidas soluções de Big Data.

A consulta de publicações recentes sobre o tema tem fornecido subsídios para formulação das respostas sobre as perguntas de pesquisa, bem como sobre ferramentas e processos para adoção de tais soluções, as quais podem maximizar a produtividade no desenvolvimento de software e implantação de Big Data.

Importante destacar que foram pesquisados e analisados principalmente estudos internacionais, de acordo com a proposta inicial de bases adotadas, assim como dos critérios.

Em estudos futuros deverá ser verificada a possibilidade de inclusão de mais artigos de produção nacional para análise de resultados que possam ser considerados mais relevantes para o caso brasileiro, incluindo uma pesquisa junto às organizações do Governo do Brasil, a qual se mostra fundamental para contribuir com produção acadêmica nacional e com o crescimento da literatura acerca do caso brasileiro.

Finalmente, a pesquisa se mostra de fundamental importância, uma vez que as dificuldades então levantadas podem ser comuns a diversos tipos de organização que pretendam implantar Ecossistemas de Big Data no Brasil, tanto públicas, nas diversas esferas de governo, quanto privadas.

#### Referências

- [1] J. Akoka, I. Comyn-Wattiau, and N. Laoufi, "Research on Big Data – A systematic mapping study," *Computer Standards & Interfaces*, vol. 54, pp. 105–115, Nov. 2017. [Online]. Available:



- <https://linkinghub.elsevier.com/retrieve/pii/S0920548917300211>
- [2] W. H. C. Almeida, L. d. A. Monteiro, A. C. de Lima, and R. R. Hazin, "Survey on Trends in Big Data: Data Management, Integration and Cloud Computing Environment," p. 7.
- [3] G.-H. Kim, S. Trimi, and J.-H. Chung, "Big-data applications in the government sector," *Communications of the ACM*, vol. 57, no. 3, pp. 78–85, Mar. 2014. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2566590.2500873>
- [4] M. d. C. Victorino, M. Shiessl, E. C. Oliveira, E. Ishikawa, M. T. de Holanda, and M. d. L. Hokama, "UMA PROPOSTA DE ECOSSISTEMA DE BIG DATA PARA A ANÁLISE DE DADOS ABERTOS GOVERNAMENTAIS CONECTADOS," p. 19, 2017.
- [5] E. Schimdt, "Every 2 days we create as much information as we did up to 2003," 2010. [Online]. Available: <http://techcrunch.com/2010/08/04/schmidt-data/>
- [6] Gantz, J., & Reinsel, D., 2011. Extracting value from chaos. IDC iView, 1142, no 2011, pp. 1-12, 2011.
- [7] Araújo Júnior, R. H., & de Sousa, R. T. B. (2016). Estudo do ecossistema de Big Data para conciliação das demandas de acesso, por meio da representação e organização da informação. *Ciência da Informação*, 45(3). [Online]. Available: <http://www.periodicos.ufpb.br/index.php/pbcib/article/view/39276>.
- [8] A. Ramos, *Infraestrutura Big Data Com Opensource*. CIENCIA MODERNA, 2015. [Online]. Available: <https://books.google.com.br/books?id=Y7HjswEACAAJ>
- [9] C. BARBIERI, *BI-business intelligence: modelagem e tecnologia*. Axcel Books, 2001. [Online]. Available: <https://books.google.com.br/books?id=HD1KAAAACAAJ>
- [10] T. Davenport, *Big data at work: dispelling the myths, uncovering the opportunities*. Harvard Business Review Press, 2014.
- [11] Márquez, F. P. G. and Lev, B., 2017. Big data management. Springer.
- [12] Inmon, W. H., & Linstedt, D., 2014. *Data architecture: a primer for the data scientist: big data, data warehouse and data vault*. Morgan Kaufmann.
- [13] D. Laney, "Application delivery strategies," META Group, Stamford, 2001.
- [14] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. Ullah Khan, "The rise of "big data" on cloud computing: Review and open research issues," *Information Systems*, vol. 47, pp. 98–115, Jan. 2015. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0306437914001288>.
- [15] O'Driscoll, A., Daugelaite, J., & Sleator, R. D., 2013. 'Big data', Hadoop and cloud computing in genomics. *Journal of biomedical informatics*, 46(5), 774-781.
- [16] Kazemi, U., 2018. *A Survey of Big Data: Challenges and Specifications*. CiiT International Journal of Software Engineering and Technology, Vol 10, No 5.
- [17] Jadhav, B., Patankar, A. B., & Jadhav, S. B. (2018). A Practical approach for integrating Big data Analytics into E-governance using hadoop. 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT). doi:10.1109/icicct.2018.8473353.
- [18] Shin, D. H., & Choi, M. J., 2015. Ecological views of big data: Perspectives and issues. *Telematics and Informatics*, 32(2), 311-320. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0736585314000665>.
- [19] Y. Demchenko, C. de Laat, and P. Membrey, "Defining architecture components of the Big Data Ecosystem," in 2014 International Conference on Collaboration Technologies and Systems (CTS). Minneapolis, MN, USA: IEEE, May 2014, pp. 104–112. [Online]. Available: <http://ieeexplore.ieee.org/document/6867550/>.
- [21] Dean, J., & Ghemawat, S., 2008. MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107-113. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=1327452.1327492>.
- [22] D. Bollier, C. M. Firestone, and others, *The promise and peril of big data*. Aspen Institute, Communications and Society Program Washington, DC, 2010.
- [23] H. E. Miller and others, "Big-data in cloud computing: A taxonomy of risks," *Information Research*, vol. 18, no. 1, pp. 18–1, 2013.
- [24] B. Klievink, B.-J. Romijn, S. Cunningham, and H. de Bruijn, "Big data in the public sector: Uncertainties and readiness," *Information Systems Frontiers*, vol. 19, no. 2, pp. 267–283, Apr. 2017. [Online]. Available: <http://link.springer.com/10.1007/s10796-016-9686-2>.
- [25] Janssen, M., & van den Hoven, J., 2015. Big and Open Linked Data (BOLD) in government: A

challenge to transparency and privacy? [Online]. Available:

<https://linkinghub.elsevier.com/retrieve/pii/S0740624X15001069>.

[26] A. M. Mariano and M. R. Santos, “Revisão da Literatura: Apresentação de uma Abordagem Integradora.” p. 18, 2017.

[27] Mariano, A. M., García Cruz, R., & Arenas Gaitán, J., 2011. Meta análises como instrumento de pesquisa: Uma revisão sistemática da bibliografia aplicada ao estudo das alianças estratégicas internacionais. Meta Analysis as a Tool of Research: A Systematic Review of Bibliography Applied Study of International Strategic Alliances. Congresso Internacional de Administração [Online]. Available:

<http://rgdoi.net/10.13140/RG.2.1.1150.1522>.

[28] Sánchez Meca, J., & López Pina, J. A., 2008. El enfoque meta-analítico de generalización de la fiabilidad. [The Meta-analytic approach of reliability generalization], *Acción Psicológica*, vol. 5, no. 2, pp. 37–64. [Online]. Available: <http://revistas.uned.es/index.php/accionpsicologica/article/view/457>.

[29] R. Khtira, B. Elasri, and M. Rhanoui, “From Data to Big Data: Moroccan Public Sector,” in *Proceedings of the 2nd international Conference on Big Data, Cloud and Applications - BDCA'17*. Tetouan, Morocco: ACM Press, 2017, pp. 1–6.

[30] F. Malomo and V. Sena, “Data Intelligence for Local Government? Assessing the Benefits and Barriers to Use of Big Data in the Public Sector: Data Intelligence for Local Government,” *Policy & Internet*, vol. 9, no. 1, pp. 7–27, Mar. 2017.

[Online]. Available: <http://doi.wiley.com/10.1002/poi3.141>.

[31] Al-Sai, Z. A., & Abualigah, L. M., 2017. Big data and E-government: A review. In *2017 8th international conference on information technology (ICIT)*, pp. 580-587, IEEE. doi: 10.1109/ICITECH.2017.8080062.

[32] E. Dogdu, M. Ozbayoglu, A. Yazici, and Z. Karakaya, 2018. Perceptions, Expectations and Implementations of Big Data in Public Sector, 2018 3Rd Int. Conf. Comput. Sci. Eng., pp. 615–620.

**APÊNDICE B – Questionário Aplicado****ANÁLISE DO IMPACTO DAS BOAS PRÁTICAS DE GOVERNANÇA DE DADOS  
ALIADO A BIG DATA NA ADMINISTRAÇÃO PÚBLICA FEDERAL**

Meu nome é Fernando Gonçalves de Oliveira, sou aluno do Mestrado Profissional em Engenharia de Software no CESAR SCHOOL sob a Orientação do Professor Felipe Silva Ferraz.

A pesquisa tem como objetivo analisar o impacto das boas práticas de Governança de Dados aliado a Big Data na Administração Pública Federal.

Sua participação é muito importante e em última análise contribui com o aprimoramento das aplicações de tecnologias da informação no âmbito da Administração Pública Brasileira.

Os resultados serão divulgados sempre em conjunto, sendo assegurado o sigilo concernente a organizações e acerca de respostas individuais.

Caso tenha interesse nos resultados ou qualquer dúvida sobre o trabalho, entre em contato pelo e-mail: fernando.infinite@gmail.com

Agradecemos pela sua importante participação nesta pesquisa!

---

1. Seu cargo na Administração Pública Federal?

- a) Servidor Público
- b) Terceirizado
- c) Outros:

2. Em qual poder trabalha?

- a) Executivo
- b) Executivo (Estatais)
- c) Judiciário
- d) Legislativo

3. Em qual órgão você trabalha?

Sua resposta

4. Qual o seu tempo de experiência em TI?

- a) 0 a 5 anos
- b) 6 a 10 anos
- c) 11 a 15 anos
- d) Mais de 15

5. Atuando em qual Estado?  
(lista dos estados)
6. Você acredita estar familiarizado com as definições e termos correlatos ao Big Data?
  - a) Sim
  - b) Não
7. Qual sua experiência com o assunto Big Data?
  - a) 0 a 5 anos
  - b) 6 a 10 anos
  - c) 11 a 15 anos
  - d) Mais de 15
8. Em seu entendimento o termo Big Data depende da volumetria de dados que órgão guarda?
  - a) Sim
  - b) Não
9. Seu órgão possui ambiente de Big Data?
  - a) Sim
  - b) Não
10. Qual a volumetria de armazenamento do órgão?
  - a) Gigabyte (GB)
  - b) Terabyte (TB)
  - c) Petabyte (PB)
  - d) Exabyte (EB)
  - e) Não sei
11. Qual a projeção de crescimento de banco de dados para 2020?
  - a) 1 a 100 Gigabyte (GB)
  - b) 1 a 100 Terabyte (TB)
  - c) 1 a 100 Petabyte (PB)
  - d) 1 a 100 Exabyte (EB)
  - e) Não sei
12. Você considera que após o crescimento projetado o seu órgão vai passar a ter um ambiente de Big Data?
  - a) Sim
  - b) Não
13. Entre os termos a seguir, quais você conhece e sabe diferenciar dos demais
  - a) MapReduce
  - b) Hadoop

- c) Business Intelligence
  - d) Data Mining
  - e) Data Science
14. Você considera que todos os termos acima possuem forte relação com o tema de Big Data?
- a) Sim
  - b) Não
  - c) A maioria SIM
  - d) A maioria não
  - e) Não tenho como opinar com segurança
15. Quais os tipos de armazenamento de dados do órgão?
- a) Servidores
  - b) Dispositivos de Storage
  - c) Cloud
  - d) Híbridos
  - e) Outros
16. Possui dados não estruturados?
- a) Sim
  - b) Não
17. Órgão possui acesso a um grande volume de dados para análise, tanto estruturado como não estruturado, numa escala de 0 a 5?
- 0 a 5
18. Existe algum problema de negócio que poderia ser resolvido com Big Data?
- a) Sim
  - b) Não
19. Existe a perspectiva de implantar Big Data no órgão, numa escala de 0 a 5?
- 0 a 5
20. Sua organização já possui projetos de Big Data?
- a) Sim
  - b) Não
21. O que mais tem motivado o uso de Big Data no seu órgão público?
- a) Necessidade da missão institucional
  - b) Transparência
  - c) Economia de recursos públicos
  - d) Volume de dados
  - e) Avanço tecnológico
  - f) Outros

22. Quais foram as dificuldades encontradas na implementação do ambiente de Big Data:
- a) Dificuldades políticas (decisão superior)
  - b) Dificuldades técnicas – domínio de tecnologias
  - c) Dificuldades financeiras
  - d) Dificuldades de alinhamento estratégico com a missão ou políticas do órgão
  - e) Ainda não tenho ambiente de Big Data
  - f) Não consigo identificar
  - g) Outros
23. Quais foram os impactos com a implantação de Big Data:
- a) Melhoria do processo de tomada decisão
  - b) Melhoria da divulgação de informações – transparência
  - c) Melhoria da eficiência no uso de recursos
  - d) Maior participação do cidadão
  - e) Melhoria do gerenciamento das informações
  - f) Ainda não tenho ambiente de Big Data
  - g) Não consigo identificar
24. Quais os impactos econômicos da adoção do Big Data em seu órgão?
- a) Não houve nenhum impacto econômico
  - b) Houve aumento de despesa em virtude dos projetos de Big Data
  - c) Houve redução de despesas em virtude dos projetos de Big Data
  - d) Houve melhoria de eficiência no gasto de recursos
  - e) Houve piora da eficiência no gasto de recursos por dificuldades de uso das tecnologias de Big Data
  - f) Não foi possível medir nenhum resultado econômico ou de despesas ou não há projetos de Big Data implementados
  - g) Não consigo identificar
  - h) outros
25. Quais as oportunidades de implementação do Big Data no seu órgão?
- a) Necessidades da missão institucional
  - b) Transparência
  - c) Economia de recursos públicos
  - d) Participação do cidadão
  - e) Requisitos tecnológicos
  - f) Volumes de dados
  - g) Não há oportunidades identificadas
  - h) outras
26. Quais os impactos que práticas de Big Data podem/poderiam trazer ao desenvolvimento de soluções em seu órgão
- a) Impactos financeiros – redução ou aumento de eficiência e eficácia

- b) Impactos tecnológicos
  - c) Impactos institucionais – missão institucional
  - d) Impactos na participação do cidadão
  - e) Não há impactos identificados até o momento
  - f) outras
27. Os impactos citados na questão anterior são mais positivos que negativos em seu órgão?
- a) Positivo
  - b) Negativo
28. A sua resposta à questão anterior envolveu impactos existentes e comprovados (Sim) ou impactos percebidos pelo conhecimento acerca de Big Data conhecido pelo respondente do questionário (Não)?
- a) Sim
  - b) Não
29. O órgão compartilha conhecimentos e aplicações de Big Data entre todos os setores do órgão, numa escala de 0 a 5?
- 0 a 5
30. A equipe de TI e profissionais de dados compartilha conhecimentos entre todos os setores da organização, numa escala de 0 a 5?
- 0 a 5
31. No ecossistema de Big Data é explorado ou adotado serviços em nuvem para processar dados ou conduzir o analytics, numa escala de 0 a 5?
- 0 a 5
32. Quantos profissionais de cientistas de dados o órgão possui?
- Terceirizados: 0 a 5 | 6 a 10 | 11 a 20 | 21 a 30 | 31 a 40 | 41 a 50
- Concursados: 0 a 5 | 6 a 10 | 11 a 20 | 21 a 30 | 31 a 40 | 41 a 50
33. Deseja se identificar, deixe seu nome e contato:
- Sua resposta

**APÊNDICE C – Questionário Aplicado – Grupo Focal****AVALIAÇÃO DE PRONTIDÃO PARA BIG DATA**

Indique seu grau de concordância em relação às seguintes afirmações referente a Avaliação de Prontidão para Big Data.

Escala Likert

1. Discordo totalmente
2. Discordo um Pouco
3. Não concordo nem discordo
4. Concordo um pouco
5. Concordo totalmente

**DADOS**

1. No órgão, temos acesso a grandes volumes de dados não estruturados ou os dados chegam em tempo real e em grande velocidade para análise.
2. No órgão, agrupamos dados de várias fontes internas em um Data Warehouse ou Datamart para facilitar o acesso.
3. No órgão, mantemos definições e padrões consistentes nos dados que usamos para análise.
4. No órgão, usuários, tomadores de decisão e desenvolvedores de soluções para áreas do órgão confiam na qualidade de nossos dados.

**NEGÓCIO DO ÓRGÃO**

5. Empregamos institucionalmente uma combinação de abordagem de Big Data e Governança de Dados para atingir as metas do órgão.
6. A nossa gestão assegura que as unidades de negócio e as áreas do órgão trabalhem para estabelecer as prioridades da organização relativas ao Big Data e Governança de Dados.
7. Capacitamos nossa equipe de TI e profissionais de análise de dados para fomentar a aprendizagem e compartilhar conhecimentos entre todos os setores da organização.



8. As nossas iniciativas e infraestrutura de Big Data e Governança de Dados recebem aportes e recursos adequados para desenvolver as competências e recursos dos quais precisamos.
9. Colaboramos com parceiros de áreas, com Órgãos para os quais prestamos suporte e outros integrantes do nosso ecossistema de negócios para compartilhar conteúdo e aplicações de Big Data.

## **LIDERANÇA**

10. Os dirigentes e líderes que estão na instituição há mais tempo ponderam regularmente sobre oportunidades que o Big Data e a Governança de Dados oferecem ao nosso negócio.
11. Os dirigentes e chefes encorajam as unidades e seus líderes a incorporar Big Data e Governança de Dados a seus processos decisórios e de negócios.
12. Os dirigentes e líderes que estão na instituição há mais tempo utilizam o Big Data e a Governança de Dados para orientar tanto suas decisões estratégicas quanto táticas.
13. Os gestores das áreas do Órgão utilizam o Big Data e o Analytics para orientar suas decisões.
14. O processo de priorização e alocação de nossos ativos de Big Data (dados, pessoal, software e hardware) é orientado e examinado pelo núcleo de altos dirigentes.

## **METAS**

15. Priorizamos nossas iniciativas de Big Data e Governança de Dados de acordo com as oportunidades que geram mais valor para nos destacarmos em termos de resultados institucionais.
16. Consideramos novos serviços e funcionalidades baseados em Big Data como um aspecto do nosso processo de inovação.
17. Identificamos oportunidades internas para Big Data e Governança de Dados avaliando nossos processos, estratégias e os ambientes onde a instituição atua.

18. Conduzimos regularmente experimentos baseados em dados para coletar dados sobre o que está – ou não – se revelando eficaz no nosso negócio.
19. Avaliamos as decisões que tomamos com apoio do Analytics e dos dados para verificar se novas fontes de dados não estruturados poderiam proporcionar modelos melhores.

## **TECNOLOGIA**

20. Temos explorado ou adotamos abordagens de computação paralela (por exemplo: Hadoop, MPI, Cuda, OpenCL e técnicas de Dataflow) para o processamento de Big Data.
21. Temos um grau elevado de maturidade no uso de visualização de dados para esclarecer questões ou decisões de negócio.
22. Temos explorado ou adotado serviços em nuvem para processar dados ou conduzir a Governança de Dados.
23. Temos explorado ou adotado software de código aberto para o Big Data e o Analytics.
24. Temos explorado ou adotado ferramentas para processar dados não estruturados, como texto, vídeo ou imagens.

## **EQUIPE DE TI E ANALISTAS DE DADOS**

25. Temos um número suficiente de potenciais cientistas de dados e profissionais de Analytics competentes para atingir nossos objetivos analíticos.
26. Nossos profissionais de tecnologias e de Analytics atuam como consultores de confiança aos nossos gestores, ajudando-os em tomadas de decisão importantes e na inovação orientada a dados.
27. Nossos profissionais de tecnologias e de Governança de Dados conhecem as disciplinas e os processos de negócio aos quais o Big Data e Analytics estão sendo ou podem ser aplicados.

28. Nossos profissionais de tecnologias e profissionais de gerenciamento de dados trabalham com eficácia em equipe na condução de projetos de Big Data, Analytics e Governança de Dados.
29. Temos programas (internos ou em parceria com organizações externas) para desenvolver ciência de dados e habilidades analíticas em nossos profissionais.

**APÊNDICE D – Entrevista Grupo Focal****AVALIAÇÃO DE PRONTIDÃO PARA BIG DATA**

1. Quais as percepções sobre as perguntas apresentadas no contexto de DADOS?
2. Quais as percepções sobre as perguntas apresentadas no contexto de NEGÓCIO DO ÓRGÃO?
3. Quais as percepções sobre as perguntas apresentadas no contexto de LIDERANÇA?
4. Quais as percepções sobre as perguntas apresentadas no contexto de METAS?
5. Quais as percepções sobre as perguntas apresentadas no contexto de TECNOLOGIA?
6. Quais as percepções sobre as perguntas apresentadas no contexto de EQUIPE DE TI E ANALISTAS DE DADOS?