

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

INSTITUTO COPPEAD DE ADMINISTRAÇÃO

ANTONIO MARQUES FERREIRA NETO

FATORES RELEVANTES NA ADOÇÃO  
E USO DE *BIG DATA* NA PREFEITURA  
DO RIO DE JANEIRO: O CASO DO  
P3NS4 – CASA DE IDÉIAS

**RIO DEJANEIRO**

**2015**

ANTONIO MARQUES FERREIRA NETO

**FATORES RELEVANTES NA ADOÇÃO E USO DE *BIG DATA* NA  
PREFEITURA DO RIO DE JANEIRO: O CASO DO P3NS4 – CASA  
DE IDÉIAS**

Dissertação de Mestrado apresentada ao  
Programa de Pós-Graduação em  
Administração, Instituto COPPEAD de  
Administração, Universidade Federal do  
Rio de Janeiro, como parte dos requisitos  
necessários à obtenção do título de Mestre  
em Administração

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Elaine Tavares

**RIO DEJANEIRO**

**2015**

ANTONIO MARQUES FERREIRA NETO

**FATORES RELEVANTES NA ADOÇÃO E USO DE *BIG DATA* NA  
PREFEITURA DO RIO DE JANEIRO: O CASO DO P3NS4 – CASA  
DE IDÉIAS**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Administração, Instituto COPPEAD de Administração, Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Administração

Orientadora: Prof<sup>ª</sup>. Dr<sup>ª</sup>. Elaine Tavares

Pesquisa realizada com apoio da Capes e da EMC.

**RIO DEJANEIRO**

**2015**

## CIP - Catalogação na Publicação

FF383f Ferreira Neto, Antonio Marques  
Fatores Relevantes na Adoção e Uso de Big Data  
na Prefeitura do Rio de Janeiro: O Caso do P3NS4  
Casa de Ideias / Antonio Marques Ferreira Neto.  
- Rio de Janeiro, 2015.  
198 f.

Orientadora: Elaine Tavares.  
Dissertação (mestrado) - Universidade Federal  
do Rio de Janeiro, Instituto COPPEAD de  
Administração, Programa de Pós-Graduação em  
Administração, 2015.

1. Big Data. 2. Fatores Relevantes. 3. Adoção e  
Uso de Big Data. 4. PENSA. 5. Data Analytics. I.  
Tavares, Elaine, orient. II. Título.

Elaborado pelo Sistema de Geração Automática da UFRJ com os  
dados fornecidos pelo(a) autor(a).

*Eu dedico essa dissertação a todos àqueles que algum dia  
Ou duvidaram de si mesmos  
Ou tiveram que superar algum grande desafio  
Ou simplesmente fizeram algo que não sabiam ser possível  
Vocês me inspiraram e iluminaram.*

## AGRADECIMENTOS

Eu consegui. Mesmo quando nem mesmo eu acreditava, eu consegui. Terminei a dissertação e o mestrado. E terminei quando até eu começava a duvidar de que isso seria possível. E talvez por isso essa seja a parte mais difícil de escrever em toda a dissertação. Não porque eu não tenho a quem agradecer. E sim porque eu tenho que agradecer a todos que contribuíram de alguma forma para que essa dissertação ficasse pronta e, ainda assim, ser breve e justo.

---

*“A sola do pé conhece toda a sujeira da estrada” (Provérbio Africano).*

Somente cada um sabe o que e por onde passou para chegar até aonde chegou. O mestrado não é uma escolha fácil. Convencer-se todos os dias de que ele vale a pena é difícil. E quando parece que as tarefas tornar-se-ão mais fáceis, aquele momento em que nos resta apenas a dissertação para fazer, nele é que residem as maiores dificuldades. E é, por isso, que eu agradeço aos meus entes queridos e amados que tanto me apoiaram e deram força. Em especial, à minha namorada, Giselle, que sentiu quando eu fraquejei e não me deixou desistir. Às minhas mãe e irmã, Gina e Olivia, que me lembraram de minhas obrigações e me cobraram pelos resultados – agradeço às três conjuntamente, inclusive por terem lido a minha dissertação sem ter necessidade ou sequer curiosidade. E, por fim – mas não menos importante –, agradeço às minhas pequenas cachorrinhas, Janis e Morena, que me ensinaram o que é amor incondicional; e que me mostraram que, às vezes, tudo o que precisamos é de um passeio e/ou de arremessar uma bolinha, pois com ela vão todos os nossos problemas. Existem provas de amor maiores que essas?

---

*“A hora mais escura do dia é aquela que precede a alvorada” (Provérbio Árabe).*

Escutei certa vez que a dissertação de um mestrado é uma tarefa tão árdua como lavar um elefante. Você não tem certeza de onde é a tromba e de onde é a cauda. Você só tem certeza de que está fazendo o seu melhor e de que está fazendo sempre que pode, mas você nunca tem certeza de qual é o seu progresso. Até que você termina. Independente das dificuldades enfrentadas. E escrever a dissertação para mim foi também “lavar um elefante”. Foi difícil. Houve momentos em que duvidei que conseguiria. Também porque fui o meu pior adversário, ao me sabotar, deprimir, entristecer, não produzir. E é

por isso que também agradeço àqueles que de alguma maneira não me deixaram desistir. Aos meus amigos, que me distraíram quando eu precisava e me entenderam e compreenderam quando eu pedi. À minha orientadora, Elaine, por ter me guiado por toda a produção da dissertação, me iluminando e ensinando, fosse para a vida ou apenas para a dissertação. Aos inúmeros que contribuíram com dados, informações, artigos e entrevistas que permitiram essa dissertação ser feita.

---

*“Se cheguei até aqui foi porque me apoiei no ombro de gigantes” (Isaac Newton).*

Enfim, agradeço a todos que, de alguma forma, me deram apoio, contribuíram e acreditaram em mim nessa etapa difícil de minha vida. Todos vocês foram gigantes para mim. Agradeço acima de tudo, no entanto, àqueles que não acreditaram e duvidaram de mim. Foram deles de quem tirei a maior motivação para terminar.

---

E, por fim, agradeço a você leitor por se dispor a ler essa dissertação. Mas não só agradeço como peço um favor especial: leia-a com a atenção devida. Leia com calma as palavras e páginas para saborear ao máximo o que cada detalhe de meu aprendizado pode contribuir com a sua sabedoria.

Desde já, obrigado. E boa leitura!

*In God we trust, for anyone else bring data.*

Michael Bloomberg



## RESUMO

Em uma época em que a informação ganha cada vez mais importância e que os dados se tornam cada vez mais disponíveis, os desafios decorrentes da falta de conhecimento tornam-se cada vez menores. Em contrapartida, aumentam os desafios devido à enormidade de possibilidades e opções com as quais se deparam as organizações. Nesse contexto, surge o termo *Big Data*, representando um conjunto de tecnologias que permite que se lide com grandes conjuntos de dados, estruturados e não estruturados, com uma velocidade anteriormente inviável. O objetivo desta pesquisa foi identificar os fatores relevantes na adoção e uso de *Big Data* no PENSA - departamento criado pela Prefeitura do Rio de Janeiro para processar dados e fazer frente aos desafios de uma administração municipal complexa. Foram encontradas evidências de que algumas características são importantes para uma implementação e uso com sucesso, dentre as quais destacam-se o alinhamento estratégico da iniciativa de implementação com a visão e objetivos organizacionais, o patrocínio da alta-gestão, o envolvimento e retenção de profissionais qualificados, o método de implementação ágil e a ética. Outros fatores como gestão de expectativas, o monitoramento e a avaliação de resultados, e a segurança não se mostraram relevantes na adoção e uso de *Big Data* no caso em questão, diferindo do que é apontado pela literatura acadêmica. A gestão da expectativa parece ter dado lugar para uma cultura de experimentação, onde as potencialidades da análise de dados são reveladas durante a realização de cada projeto.

**Palavras-chave:** *Big Data*, Análise de dados, Fatores Relevantes, Implementação, Uso, PENSA, Prefeitura do Rio de Janeiro.

## **ABSTRACT**

In an age where information is becoming increasingly important and data are becoming increasingly available, the challenges arising from lack of knowledge become increasingly smaller. In contrast, challenges are increased due to new possibilities that organizations face. In this context, the term Big Data arises, representing a set of technologies that enables organizations to handle large sets of data, structured and unstructured, with an unforeseen speed. The objective of this research was to identify relevant factors in the adoption and use of Big Data at PENSA, a department created by Rio de Janeiro City Hall to process data and to meet the challenges of a complex municipal administration. Evidence was found that some features are important for a successful implementation and use, among which stand out the strategic alignment of the implementation initiative with vision and organizational goals, sponsorship of top management, involvement and retention of qualified professionals, the implementation of agile management methods and ethics. Other factors, traditionally associated to successful implementation of Business Intelligence projects, were not perceived as relevant for Big Data adoption and use. Other factors such as management of expectations, monitoring and evaluation of results and safety were not relevant in the adoption and use of Big Data in this case studied, which contradicts the academic literature. Expectations of management seems to have been replaced for a culture of experimentation, where data analysis capabilities are revealed during the achievements of each projects.

**Key words:** Big Data, Data analysis, Relevant Factors, Implementation, Use, PENSA, Rio de Janeiro City Hall

# SUMÁRIO

1. Introdução .....	15
1.1. Contextualização do Problema .....	15
1.2. Objetivo Geral .....	17
1.3. Objetivos Específicos .....	17
1.4. Relevância .....	18
1.5. Delimitações .....	20
2. Big Data .....	21
2.1. As Principais Dimensões do Conceito.....	21
2.2. Outras Dimensões do Conceito .....	23
2.3. Além da Definição Multidimensional .....	25
2.4. O Conceito de <i>Big Data</i> Adotado.....	26
2.5. <i>Big Data</i> e demais sistemas tradicionais de análise de dados .....	26
2.6. Evolução e Desenvolvimento do <i>Big Data</i> .....	28
2.6.1. Evolução Conceitual.....	29
2.6.2. Evolução Temporal da Análise de <i>Big Data</i> .....	32
2.7. Benefícios e Oportunidades.....	35
2.7.1. Benefícios e oportunidades setoriais .....	36
2.7.2. Benefícios e Oportunidades de Gestão.....	45
2.7.3. Oportunidades em Todos os Departamentos de uma Organização .....	54
3. Fatores Relevantes na Adoção e no Uso de <i>Big Data</i> .....	57
3.1. Fatores Relacionados à Estratégia .....	57
3.1.1. Alinhamento à Estratégia da Organização.....	57
3.1.2. Patrocínio da Alta Gestão .....	59
3.2. Fatores Estruturais e Culturais.....	60

3.2.1. Design e Processos Organizacionais .....	60
3.3. Fatores Relacionados aos Recursos Humanos.....	63
3.3.1. Talent Gap .....	63
3.3.2. Formação de Equipe .....	68
3.3.3. Gestão de Expectativas .....	72
3.4. Fatores Relevantes Relacionados à Gestão do Programa de Implementação .....	73
3.4.1. Comitê Gestor, Grupo de Superusuários e Campeão de Projeto .....	73
3.4.2. Comunicação .....	75
3.4.3. Programa de Implementação .....	76
3.4.4. Monitoramento e Avaliação de Performance .....	78
3.5. Fatores Relacionados à Governança da Informação.....	80
3.5.1. Ética .....	80
3.5.2. Privacidade .....	82
4. MODA: modelo no qual foi baseado o PENSEA .....	87
4.1. Antecedentes.....	87
4.1.1. O Desenvolvimento do Modelo Compstat .....	87
4.1.2. Propagação do Modelo Compstat para outras agências .....	88
4.1.3. Desenvolvimento do Modelo GOV 2.0.....	89
4.2. MODA – Mayor’s Office of Data Analytics .....	91
4.2.1. Premissas e Objetivos.....	91
4.2.2. Equipe e Estrutura .....	93
Lições do MODA para futuras equipes de análise de dados no setor público .....	94
5. Método.....	95
5.1. Tipo de pesquisa .....	95
5.2. O Caso selecionado .....	96
5.3. Coleta e Análise de dados.....	96

5.4. PENSA – Sala de Ideias .....	98
6. Análise de Dados .....	102
6.1. Fatores Organizacionais .....	102
6.1.1. Alinhamento à Estratégia.....	102
6.1.2. Patrocínio da Alta Gestão .....	105
6.1.3. Alinhamento aos processos organizacionais .....	107
6.2. Fatores Relacionados aos Recursos Humanos.....	111
6.3. Fatores Relacionados à Implementação .....	120
6.3.1. Estruturas facilitadoras da Implementação.....	120
6.3.2. Programa de Implementação .....	123
6.3.3. Abordagem Faseada e Evolucionária .....	125
6.4. Fatores Relacionados à Percepção de Valor de Negócios.....	130
6.4.1. Gestão de Expectativas.....	130
6.4.2. Comunicação .....	132
6.4.3. Monitoramento e Avaliação de Performance .....	135
6.4.4. Ética .....	138
6.4.5. Privacidade e Segurança.....	141
6.5. Novos candidatos à Fatores Relevantes.....	145
6.5.1. Estrutura.....	145
6.5.2. Infraestrutura/Arquitetura de Sistema .....	147
Desafios relacionados aos dados .....	150
7. Conclusão .....	154
Bibliografia.....	160
Anexo I – Roteiro de Pesquisa .....	172
Anexo II – Lista de definições de <i>Big data</i> .....	175
Anexo III – Evolução Conceitual e Temporal de Sistemas de Análise de Dados.....	184

Anexo IV – Projetos de <i>Big Data</i> ao redor do Mundo .....	185
Anexo V – Decreto de Criação do MODA (Modelo de Nova York).....	186
Anexo VI - Modelo de Processos do MODA.....	189
Anexo VII – Decreto de Criação do Pensa.....	190
Anexo VIII – Decreto de Acesso aos Dados para o Pensa.....	192

# 1. INTRODUÇÃO

## 1.1. CONTEXTUALIZAÇÃO DO PROBLEMA

A análise de dados no contexto organizacional vem passando por grandes transformações. Se antes era necessário um programa de relacionamento e fidelidade para identificar o perfil de cada consumidor por meio de seu histórico de compras, hoje este perfil pode ser estimado por uma companhia mesmo que nenhuma compra tenha sido realizada com ela, devido ao que as pessoas divulgam online, consciente ou inconscientemente, sobre suas experiências de compra e de vida (CANARY, 2013; TOLE, 2013).

A quantidade de informação disponível é abundante e cresce cada vez mais. Em 2012, por exemplo, aproximadamente 2,4 *exabytes* de dados foram criados diariamente (MCAFEE e BRYNJOLFSSON, 2012). A previsão é que em 2020 será produzido mais de 40.000 *exabytes*, mais de 5.200 gigabytes para cada habitante do planeta (GANTZ e REINSEL, 2012). Hoje mais dados se tornam disponíveis na Internet a cada segundo do que toda a capacidade de armazenamento da internet há vinte anos atrás. O número de dispositivos em rede de dados ultrapassou a população global em 2011 e os analistas estimam que cinquenta bilhões de sensores serão conectados à internet em 2025 (DAVENPORT, 2014). A quantidade de dados trafegados na internet crescerá em velocidade igual ou superior: a Cisco estima que, em 2016, 130 *exabytes* de dados vão trafegar através da Internet a cada ano (FITCHARD, 2012). Ao mesmo tempo, é esperado na mesma proporção uma evolução tecnológica no processamento, organização e gerenciamento do alto volume de dados (SILVA e CAMPOS, 2014).

Não há dúvidas que estamos gerando e consumidos mais dados – e *diferentes tipos de dados* – do que nunca (SIMON, 2013). Na era da informação, o desafio das organizações não é obter dados para analisar, e sim obter informações relevantes para a tomada de decisões em meio à grande quantidade de dados disponíveis. A informação auxilia no processo decisório, pois é a base da tomada de decisão (OLIVEIRA, 2002).

Foi devido a necessidade de se lidar com o grande e crescente volume de informações que surgiram os Sistemas de Informação (SI), que dão condições para as empresas reagirem às mudanças de mercado e ter processos de decisão eficazes para resolver seus problemas e superar seus desafios (BAZZOTTI e GARCIA, 2006). Em específico, os sistemas relacionados a *Business Intelligence* (BI), e mais recentemente os sistemas relacionados à *Big Data*, têm adquirido cada vez mais importância nos negócios nos últimos anos, pois as

necessidades das organizações concentram-se cada vez mais na análise de dados e na rápida tomada de decisões baseada em grandes volumes de informação (CHEN, CHIANG e STOREY, 2012).

De acordo com o *Big Data Executive Survey* (2013), uma pesquisa realizada com os executivos seniores de negócios tecnológicos da *Fortune 1000* realizada pela *NewVantage Partners*, 68% dos executivos esperavam que suas organizações investissem mais de um milhão de dólares em *Big Data* em 2013. Esta porcentagem subia de forma constante até o patamar de 88%, quando se considerava 2016 como horizonte de tempo. Investimentos superiores a dez milhões de dólares apresentavam uma tendência de crescimento ainda mais acentuada, com projeção subindo de 19% em 2013 para 50% até 2016.

Nesse contexto, o *Gartner Group* estima que os investimentos em *Big Data* no mundo foram da ordem de US\$ 28 bilhões em 2012, com projeção de US\$ 34 bilhões em 2013 (STAMFORD, 2012). De forma semelhante, a consultoria Deloitte espera enorme crescimento do *Big Data*, embora estimar precisamente o tamanho do mercado seja um desafio (SIMON, 2013).

Essas expectativas são reflexos dos benefícios esperados do uso do *Big Data*: integração de dados de diversas fontes, análise de grandes volumes de dados estruturados e não estruturados, melhora das capacitações analíticas gerais, análise de informação em tempo real, redução de custos da análise e processamento de dados, maior e melhor captura de dados, maior transparência das informações utilizadas para a tomada de decisões, personalização mais ágil de produtos e serviços, e criação de novos modelos de negócios baseados em dados (BROWN, CHUI e MANYIKA, 2011; DAVENPORT, BARTH e BEAN, 2012).

A expansão da TI no contexto organizacional criou também desafios: as rápidas mudanças em TI fazem com que o ambiente de negócios se torne ainda mais imprevisível. A habilidade das organizações para identificar informações relevantes tornou-se crucial, já que o acesso aos dados usados para gerar informação não é mais restrito. TI permite gravar, sintetizar, analisar e disseminar informação mais rápido que em qualquer outro momento da história. Os dados podem ser coletados de diferentes partes da companhia e seu ambiente externo e podem contribuir para o desenvolvimento de uma vantagem competitiva (GALLIERS e LEIDNER, 2003).

A disseminação do uso de SIs é motivada por diversos resultados de várias ordens, geralmente relacionados a ganhos de produtividades, de competências e de desempenho



organizacional (RODRIGUES, 2008). Dentre os resultados mais comuns em pesquisas na área está o fato de que a tecnologia por si só não pode solucionar os problemas e que a introdução da tecnologia resulta em mudança. O impacto da mudança tecnológica depende do por que e do como a tecnologia é usada (GALLIERS e LEIDNER, 2003).

Com *Big Data* não poderia ser diferente. *Big Data* envolve a transformação nos negócios – transformar a organização no sentido de monitorar o ambiente de negócios, em tempo real, e desenvolver modelos de previsão (SCHMARZO, 2013). Trata-se de um conjunto de soluções que pode aprimorar muito os processos de decisão gerencial, orientando as organizações para dados e com possibilidade de gerar novos modelos de negócios.

O impacto positivo do *Big Data* não se limita às organizações de capital privado. Mesmo os governos podem se beneficiar da captura e análise de grandes volumes de dados estruturados e não estruturados, por meio, por exemplo, da personalização da experiência dos cidadãos, da redução da fraude nos impostos e na seguridade social ou da melhora de áreas críticas dependentes dos investimentos públicos tais como saúde e segurança pública (VAN RIJMENAM, 2014). A esses benefícios, somam-se muitos outros, como poderá ser visto no presente trabalho.

Mediante este cenário, este trabalho foi orientado pela seguinte pergunta de pesquisa: os fatores relevantes na adoção e uso de *Big Data* são os mesmos fatores que se revelaram relevantes na adoção e uso de sistemas de análise de dados anteriores, tais como sistemas de *Business Intelligence*?

## **1.2. OBJETIVO GERAL**

O objetivo geral desta pesquisa foi identificar quais foram os fatores relevantes na adoção e o uso de *Big Data* na prefeitura municipal do Rio de Janeiro, no programa PENSA, também conhecido por “P3NS4 – Sala de Ideias”.

## **1.3. OBJETIVOS ESPECÍFICOS**

Diante do objetivo geral proposto, esta pesquisa se desdobrou nos seguintes objetivos específicos.

- a) Definir o conceito de *Big Data*, e investigar a evolução conceitual e temporal do termo;
- b) Identificar as oportunidades na adoção e uso de *Big Data* no contexto organizacional, bem como os benefícios oriundos da adoção;
- c) Identificar os fatores relevantes na adoção e uso de *Big Data* no contexto organizacional;

O primeiro objetivo foi entender o conceito de *Big Data*, evidenciando o que o torna diferente de tecnologias anteriores (TAURION, 2012), tendo em vista os diversos e distintos conceitos existentes de *Big Data*.

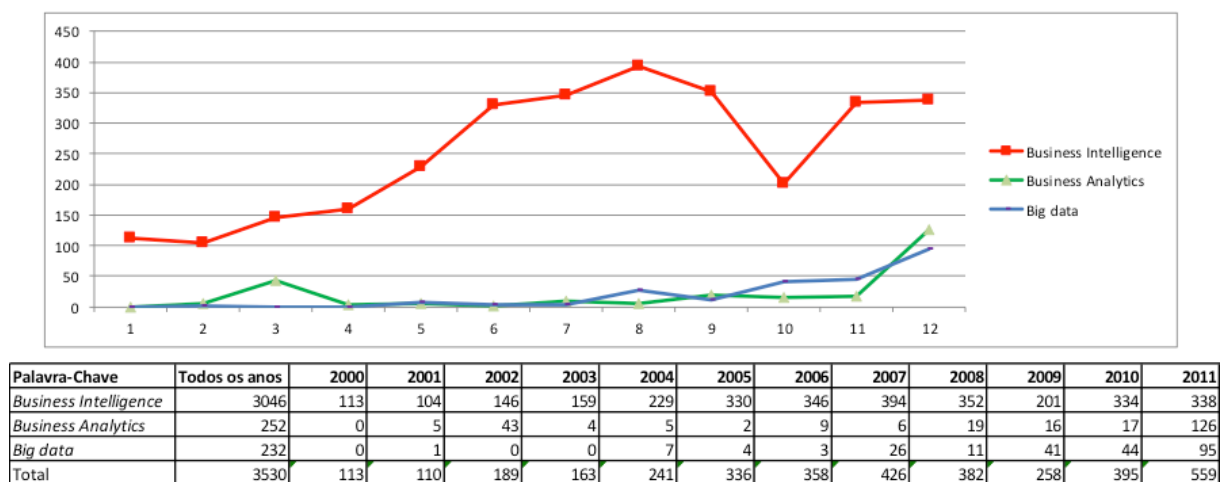
A partir de então, os demais objetivos específicos foram associados à identificação do benefícios, oportunidades e fatores relevantes na sua adoção e uso. Como se trata de um conjunto de tecnologias bem recente, recorreu-se à literatura não somente dos aspectos gerenciais de *Big Data* bem como à literatura de sistemas de informação, sobretudo sistemas *Business Intelligence* (BI). Esta opção se deu, primeiro, porque tal qual os sistemas de BI, as tecnologias associadas ao *Big Data* objetivam a melhora do processo de tomada de decisões, por meio da captura de informações e análise de dados (DAVENPORT, 2006; YEOH e KORONIUS, 2010). Segundo, porque *Big Data* tem sob sua égide muitas tecnologias, métodos e conceitos de análise que não são novos, dentre eles a maioria dos sistemas tradicionais de BI e *Data Mining* (CHEN, CHIANG e STOREY, 2012; OHLHORST, 2013). Terceiro, porque a literatura já apontou que alguns fatores de sucesso para *Big Data* e projetos de análise de dados são similares àqueles de implementações de sistemas de BI (EVERS, 2014).

## 1.4. RELEVÂNCIA

Com o aumento do volume de dados na web, em dispositivos móveis, em sensores e em diversas outras fontes de dados disponíveis, e com o desenvolvimento das tecnologias e ferramentas analíticas, tornou-se possível produzir *insights* para os processos de negócios a partir de conteúdos extremamente detalhados, contextualizados, ricos e relevantes para os negócios (CHEN, CHIANG e STOREY, 2012; SCHMARZO, 2013). Tal possibilidade atraiu a atenção da comunidade de negócios e da sociedade, que passou a elencar *Big Data* como uma das principais tendências tecnológicas da atualidade (IBM, 2011; CEARLEY e

CLAUNCH, 2012; DAVENPORT, 2014; RIBEIRO, 2014), sobretudo depois de algumas empresas terem se desenvolvido e crescido com a oferta de produtos baseados em *Big Data*, tal como a Amazon, o Google e o Facebook (OHLHORST, 2013).

O ganho de importância na comunidade de negócios logo se refletiu na comunidade acadêmica com um crescimento constante e crescente de publicações sobre Business Intelligence, Business Analytics e Big Data, tal como apresenta a Figura 1 (CHEN, CHIANG e STOREY, 2012; SILVA e CAMPOS, 2014). Silva e Campos (2014) destacam que tendência semelhante é encontrada quando a quantidade de citações ao tema é analisada.



**Figura 1:** Tendência de Publicação relacionada à BI&A entre 2000 e 2011 (Fonte: CHEN, CHIANG & STOREY, 2012)

Apesar do desenvolvimento tecnológico, uma pesquisa da *Harvard Business Review* (HBR) com quase mil leitores, em 2013, revelou que o uso de tecnologias de *Big Data* ainda apresenta diversos desafios gerenciais: apenas 6% afirmaram que a organização considerou o impacto de *Big Data* sobre as principais funções da empresa, e uma porcentagem ainda menor – de 3,5% – declarou que a organização sabia como aplicar *Big Data* para o negócio. Trata-se de um conjunto novo de tecnologias analíticas que deve ter grande impacto na tomada de decisão gerencial e pode representar mudanças de modelos mentais nas organizações, exigindo uma série de novas competências individuais e organizacionais.

Em suma, *Big Data* representa um novo paradigma, no qual “a ênfase na indústria mudou para a análise de dados e rápida tomada de decisões baseada em grandes volumes de informação” (CHEN, CHIANG e STOREY, 2012). Um conjunto de tecnologias em que a maior parte dos desafios está no nível da implementação (LABRINIDIS e JAGADISH, 2012).

Entretanto, o meio acadêmico também está apenas iniciando a discussão do tema, considerando a aplicação do Big Data nas organizações. Após uma pesquisa bibliográfica realizado no início desta pesquisa em 2014, apenas um artigo, Evers (2014), foi encontrado sobre os fatores de sucesso para aplicações de *Big Data* na perspectiva gerencial. Faz-se necessário, portanto, investigar que oportunidades, benefícios e fatores relevantes se destacam para uma implementação e uso de sucesso da tecnologia.

## 1.5. DELIMITAÇÕES

Tendo em vista o tema, a primeira delimitação dessa pesquisa é a tecnologia cujo uso será pesquisado. A presente pesquisa busca explorar os fatores relevantes na adoção e uso de *Big Data*, por ser essa uma tecnologia recente, de grande relevância para a gestão, e por ser um tema ainda pouco explorado. Outros sistemas de *analytics* não farão parte do escopo da pesquisa.

De maneira geral, vale ressaltar ainda como limitação o fato de os resultados da pesquisa terem sido obtidos num contexto específico. Isso significa que o estudo foi conduzido junto a um departamento de uma organização do setor público, cujas características podem ser bastante distintas de organizações do setor privado. Logo, os resultados encontrados são válidos considerando esse contexto, necessitando de apreciação, análise e validação em contextos diferentes.

É interessante destacar que o período temporal em que a pesquisa é realizada também faz parte do contexto. Logo, os resultados encontrados, dispostos e explicados aqui são válidos respeitando essa janela temporal – ao longo do segundo semestre de 2014 e primeiro semestre de 2015 –, tendo em vista que ao longo do tempo, com o desenvolvimento e da maturidade do *Big Data*, os fatores relevantes no seu uso podem ter mudado.

## 2. BIG DATA

*Big Data* define uma situação em que os conjuntos de dados cresceram de tal maneira que tecnologias da informação convencionais não podem efetivamente lidar com sua proporção. Em outras palavras, o conjunto de dados é tão grande que é difícil de gerir e ainda mais difícil angariar valor do mesmo (OHLHORST, 2013).

*Big Data* está relacionado aos conjuntos de dados, cujo tamanho é além da capacidade de ferramentas típicas de software de banco de dados para capturar, armazenar, gerenciar e analisar. Conforme a tecnologia avança ao longo do tempo, o tamanho dos conjuntos de dados que sejam qualificados como *Big Data* também irá aumentar (MINELLI, CHAMBERS e DHIRAJ, 2013).

A dimensão do conjunto de dados expõe apenas parte do que é compreendido como *Big Data*. Apesar de ser o atributo mais conhecido do *Big Data*, o volume não é sua única característica importante (RUSSOM, 2011; LANEY, 2001). Anos antes que o termo *Big Data* fosse cunhado, Laney (2001) definiu três dimensões primárias do que ele chamou de dilúvio de dados: volume (a quantidade crescente de dados), variedade (o aumento da gama de tipos de dados e fontes), e velocidade (o aumento da velocidade de dados). Os três v's de Laney se tornaram referência quando se trata de *Big Data* (SIMON, 2013).

De acordo com a Gartner Inc., *Big Data* são ativos de informação de grande volume, velocidade e variedade que exigem formas custo-efetivas e inovadoras de processamento de informações para uma melhor percepção e tomada de decisões (CEARLEY e CLAUNCH, 2012). Volume, velocidade e variedade da informação constituem a primeira parte da definição de *Big Data* da Gartner – as três dimensões diretrizes do *Big Data* (SICULAR, 2013).

### 2.1. AS PRINCIPAIS DIMENSÕES DO CONCEITO

Conforme apresentado no item anterior, o mercado adotou uma definição evolutiva segundo a qual Big data que está definida por três dimensões: Volume, Variedade e Velocidade (MINELLI, CHAMBERS, & DHIRAJ, 2013).

De acordo com um estudo recente da McKinsey Global Institute, as organizações capturam trilhões de bytes de informações sobre seus clientes, fornecedores e operações através de

sistemas digitais. Milhões de sensores em rede embutidos em telefones celulares, automóveis e outros produtos estão continuamente detectando, criando e comunicando dados. O resultado é uma projeção de crescimento anual do volume de dados gerados de 40%. Como nota o estudo, 15 dos 17 setores da economia dos EUA já "tem mais dados armazenados por empresa do que a Biblioteca do Congresso dos EUA" (MINELLI, CHAMBERS e DHIRAJ, 2013).

Há menos de 10 anos atrás, usava-se muito o termo *gigabytes* e rapidamente mudou-se para *terabytes*. Hoje, as empresas são inundadas com os dados, acumulando facilmente *terabytes* e até mesmo *petabytes* e *exabytes* de informações, numa tendência crescente (OHLHORST, 2013; SILVA e CAMPOS, 2014).

É importante observar que a definição de volume pode variar de acordo com o setor, dependendo dos tipos de ferramentas de software disponíveis e dos tamanhos de conjuntos de dados usuais do setor. Com essas ressalvas, a definição de *Big Data* em muitos setores hoje varia de algumas dezenas de *terabytes* para múltiplos *petabytes* (MINELLI, CHAMBERS e DHIRAJ, 2013).

Dentre os principais efeitos destes volumes estão seu impacto nos processos organizacionais e suas influências nos métodos preditivos e estatísticos (MINELLI, CHAMBERS e DHIRAJ, 2013).

Volume não é a única dimensão importante do *Big Data*. A velocidade também o é. E por velocidade de dados entende-se a velocidade à qual os dados são criados, inseridos, acumulados e processados (MINELLI, CHAMBERS e DHIRAJ, 2013).

Em um mercado altamente competitivo, encontrar novas maneiras de interpretar os dados e processá-los de maneira mais rápida tem se mostrado uma vantagem competitiva crítica e está direcionando *Big Data analytics* para novos processos (OHLHORST, 2013).

A velocidade como característica de *Big Data* também se refere às taxas de transferência, à vinculação de conjuntos de dados que estão vindo com velocidades diferentes e a mudança repentina das quantidades de informação transferidas quando comparadas ao ritmo habitual (SICULAR, 2013). Em outros termos, velocidade pode significar uma taxa de transferência por segundo tão alta que os sistemas não têm como armazenar todos os dados transferidos e, por isso, as organizações são obrigadas a escolher quais dados capturar e quais dados descartar (XEXÉO, 2013).

Para muitas aplicações, como, por exemplo, as seguradoras, a velocidade é ainda mais importante do que o volume. Informações em tempo real e os padrões de como os dados se

relacionam tornam possível para a companhia se tornar muito mais ágil que seus competidores (MCAFEE e BRYNJOLFSSON, 2012).

Além da quantidade e da velocidade, a variedade é outro conceito presente. Muitas das mais importantes fontes de *Big Data* são relativamente novas, o que torna os bancos de dados estruturados tradicionais, que armazenam informações corporativas, inadequados para o armazenamento e processamento destes novos tipos de dados (MCAFEE e BRYNJOLFSSON, 2012).

*Big Data* se estende para além de dados estruturados para incluir dados não estruturados de todas as variedades (OHLHORST, 2013). *Big Data* toma a forma de mensagens, atualizações, e imagens postadas em redes sociais, leituras de sensores, sinais de GPS de celulares, dentre outros. Além disso, existem muito mais tipos de dados do que aqueles que se podem analisar por um sistema de gestão tradicional (MCAFEE e BRYNJOLFSSON, 2012).

Variedade de dados é, então, a quantidade de tipos de dados utilizados, considerando não apenas dados estruturados, mas também dados semiestruturados e não estruturados. Dados não estruturados são, basicamente, informações que, ou não tem um modelo de dados pré-definido e/ou não se encaixam bem em um banco de dados relacional. O termo semiestruturado é usado para descrever dados estruturados que não se encaixam em uma estrutura formal de modelos de dados. No entanto, os dados semiestruturados contém marcas que separam elementos semânticos, o que inclui a capacidade de impor hierarquias dentro dos dados (MINELLI, CHAMBERS e DHIRAJ, 2013).

O verdadeiro desafio do *Big Data* é encontrar a melhor maneira de lidar com as fontes de dados variadas e ainda atender os objetivos do processo analítico. Esta é uma abordagem mais experiente que integra hardware, software e procedimentos em um processo administrável que proporciona resultados dentro de um prazo aceitável (OHLHORST, 2013).

## **2.2. OUTRAS DIMENSÕES DO CONCEITO**

Embora constituam uma definição compreensiva do que está e do que não está relacionado à *Big Data*, volume, variedade e velocidade não são os únicos traços característicos (LANEY, 2001; RUSSOM, 2011). Como consequência, alguns definiram *Big Data* em termos de três V's (volume, variedade e velocidade); outros têm acrescentado ainda mais V's (DAVENPORT, 2014). Outros autores, como, por exemplo, Brown, Chui e Manyika (2011),

Tankard (2012), Taurion (2012), e Canary (2013) têm destacado outras duas dimensões que complementam a relevância do fenômeno. São elas: valor e veracidade.

Apesar de haver algum nível de consenso sobre a importância da dimensão valor, a definição de valor varia entre os diversos autores. Como sintetizaram Maçada e Canary (2014), valor pode ser definido como resultado agregado a partir das análises das informações, como qualidade das informações ou como valor financeiro a partir de dados de qualidade.

Com relação ao valor, Chen, Chiang e Storey (2012) afirmam que as análises críticas de dados podem ajudar as empresas a melhor entender seus negócios. Brown, Chui e Manyika (2011) e Tankard (2012) compartilham de opiniões semelhantes envolvendo a dimensão valor como o resultado agregado obtido da análise dos dados coletados (CANARY, 2013).

Brown, Chui e Manyika (2011) destacam a confiança e a experiência para a extração de valor na análise rigorosa de dados do *Big Data*. Tankard (2012) destaca a clareza como uma característica importante para atribuir valor às informações.

Webber, Otto e Osterle (2009) ressaltam que a qualidade dos dados exige “exatidão, integridade, consistência e relevância”, e ainda destacam a dependência da qualidade de informação percebida pelo usuário nas suas necessidades.

Valor é importante porque é absolutamente necessário que a organização que implementa projetos de *Big Data* obtenha retorno destes investimentos (BEATH, BECERRA-FERNANDEZ, *et al.*, 2012; TAURION, 2012). Beath e seus colegas (2012) ressaltam que ao realizar investimentos para o *Big Data*, as empresas focam em armazenar, proteger e acessar essa grande quantidade de dados, enquanto que deveriam estar focadas também em buscar o valor da informação, concentrando-se nas oportunidades para aumentar o valor do negócio.

O valor muitas vezes é impulsionado pela busca de vantagem competitiva, encorajando as organizações a transformar seus grandes repositórios de dados corporativos e externos para descobrir tendências, estatísticas e outras informações úteis, para ajudá-los a decidir sobre o seu próximo movimento (OHLHORST, 2013).

A veracidade está intimamente associada ao valor. As enormes quantidades de dados recolhidos para fins de *Big Data* podem levar a erros estatísticos e má interpretação da informação recolhida. A pureza e autenticidade das informações são fundamentais para o valor (TAURION, 2012; OHLHORST, 2013).



Pipino, Yang e Wang (2002), por exemplo, consideram a “credibilidade” como outra dimensão da qualidade da informação. Informações verdadeiras têm relação com informações exatas, íntegras, consistentes e relevantes - características da qualidade da informação -, podendo, então, ser utilizadas pelos gestores ou responsáveis para responder aos desafios estratégicos e operacionais (WEBER, OTTO e OSTERLE, 2009). Uma tomada de decisão não pode ser embasada em dados duvidosos. A veracidade garante, então, a confiabilidade e credibilidade dos dados.

### **2.3. ALÉM DA DEFINIÇÃO MULTIDIMENSIONAL**

Além das diversas dimensões representadas por palavras iniciadas com a letra v, a tecnologia também é um aspecto importante da definição de *Big Data*.

É necessário pensar sobre as capacidades das tecnologias para armazenar e processar dados não estruturados; sobre a ligação de dados de diversos tipos, origens e taxas de variação; e sobre a realização de uma análise abrangente (SICULAR, 2013). Em outras palavras, é necessário pensar sobre o conjunto de tecnologias – que não se reduzem a apenas uma – que permitam analisar as dimensões diretrizes dos conjuntos de dados.

A tecnologia que melhor representa *Big Data* é o algoritmo conhecido como *MapReduce* e sua implementação mais conhecida é a plataforma *Hadoop*. Pesquisadores do Google publicaram em 2008 um artigo descrevendo a tecnologia desenvolvida pela companhia para executar buscas na Web com dados armazenados em muitos computadores diferentes (DEAN e GHEMAWAT, 2008). O nome dado ao algoritmo, *MapReduce*, se refere às suas duas operações principais: "mapear" uma consulta em diversos computadores, e "reduzir" – isto é, combinar – os resultados para resultar numa resposta consolidada para o usuário final. Pesquisadores do *Yahoo!Labs* se interessaram pela tecnologia e criaram uma implementação de código aberto, que chamaram de *Hadoop* (SHVACHKO, KUANG, *et al.*, 2010).

Importante é destacar que *Big Data* não está relacionada a soluções de baixo custo, e sim a soluções tecnológicas custo-efetivas dos problemas relacionados à análise de dados (SICULAR, 2013). As tecnologias de *Big Data* associam abordagens de processamento inovadoras por meio de servidores de baixo custo com uma grande variedade cada vez mais robusta de código-fonte aberto. Isso ocasionou uma mudança na equação de custo-benefício de processamento da informação, pois as tecnologias trazem benefícios muito maiores que os

custos de obtê-los (CEARLEY e CLAUNCH, 2012). Em suma, além do próprio conjunto de dados em si, as tecnologias utilizadas para a armazenagem e análise dos dados também são importantes.

## **2.4. O CONCEITO DE *BIG DATA* ADOTADO**

Nesta pesquisa, *Big Data* será entendido como um conjunto de tecnologias capazes de armazenar, limpar, tratar, processar, analisar e extrair valor de conjuntos de dados que são difíceis de processar com tecnologias tradicionais (e que podem ser definidos por um ou mais V's do conjunto multidimensional de V's), bem como o próprio conjunto de dados que demanda essas tecnologias para serem analisados.

Uma tabela com as referências utilizadas nesta pesquisa para o entendimento do conceito de *Big Data* é apresentada no anexo II.

## **2.5. *BIG DATA* E DEMAIS SISTEMAS TRADICIONAIS DE ANÁLISE DE DADOS**

Algumas pessoas enxergam *Big Data* como meramente outro nome para *Business Intelligence* (BI), sobretudo porque é possível observar diversas ferramentas tradicionais de mineração de dados que prometem derivar valor de grandes conjuntos de dados. No entanto, isso não significa que BI – e outras ferramentas de análise de dados – e *Big Data* são a mesma coisa (SIMON, 2013).

Parte da confusão deriva do fato de muitas das tecnologias, análises e conceitos utilizados para definir o valor das fontes de dados pelo *Big Data* não serem novos. Melhor definidas como categorias de análise, estas tecnologias incluem sistemas de *Data Mining*, aplicações estatísticas, técnicas de modelagem de dados e técnicas de análise preditiva (OHLHORST, 2013).

Muitas dessas técnicas provêm de Estatística e foram usadas em painéis de BI e *scorecards* (MINELLI, CHAMBERS e DHIRAJ, 2013). Muitas ferramentas de mineração de dados estão se expandindo e evoluindo, de modo a comportar conjuntos cada vez maiores de dados (SIMON, 2013). Além disto, muitos fornecedores de TI e provedores de soluções usam o

termo *Big Data* como sinônimo de análise de dados mais inteligente e perspicaz (DAVENPORT e PATIL, 2012).

A figura a seguir, extraída de Minelli, Chambers e Dhiraj (2013), evidencia a diferença entre os sistemas tradicionais de BI e sistemas de análises de dados avançados, muitos dos quais se encaixam sob a égide de *Big Data*.

Análise por Linguagem de Consulta Estruturada (SQL)	Análise descritiva	Mineração de dados	Análise preditiva	Simulação	Otimização
<ul style="list-style-type: none"> <li>Contagem</li> <li>Média</li> <li>OLAP</li> </ul>	<ul style="list-style-type: none"> <li>Distribuição univariada</li> <li>Tendência Central</li> <li>Dispersão</li> </ul>	<ul style="list-style-type: none"> <li>Regras de Associação</li> <li>Clusterização</li> <li>Extração de características</li> </ul>	<ul style="list-style-type: none"> <li>Classificação</li> <li>Regressão</li> <li>Previsão espacial</li> <li>Machine learning</li> <li>Análise de textos</li> </ul>	<ul style="list-style-type: none"> <li>Monte Carlo</li> <li>Modelagem baseada no agente</li> <li>Modelagem de eventos discretos</li> </ul>	<ul style="list-style-type: none"> <li>Otimização Linear</li> <li>Otimização não linear</li> </ul>

**Figura 2:** Evolução das técnicas de análise de dados (Fonte: MINELLI, CHAMBERS E DHIRAJ, 2013)

O *Big Data* pode ser entendido como uma evolução do BI, que não tem a pretensão de substituir o conceito. Brynjolfsson (2012) afirma que soluções de *Big Data* possuem um potencial ainda maior de beneficiar as empresas do que as soluções analíticas tradicionais, as quais já são utilizadas e conhecidas por abrir novas oportunidades e oferecer vantagens competitivas para aquelas que as utilizam.

A tabela a seguir tem como objetivo apresentar as principais diferenças entre *Big Data* e as análises de dados tradicionais (DAVENPORT, 2014).

	Big Data	Sistemas Tradicionais de Análise de Dados (BI)
Tipo de Dados	Formatos desestruturados	Formatados em linhas e colunas
Volume de Dados	De 100 terabytes a petabytes	Dezenas de terabytes ou menos
Fluxo de Dados	Fluxo contínuo de dados	Conjunto estático de dados
Método de Análise	<i>Machine Learning</i>	Baseado em Hipóteses
Propósito primário	Produtos baseados em dados	Suporte e serviços de decisão interna

**Tabela 1:** Definição de *Big Data* (DAVENPORT, 2014)

Novo & Neves (2013) e McAfee e Brynjolfsson (2012) defendem que as características que definem multidimensionalmente o *Big Data* são justamente aquelas que o diferenciam de sistemas tradicionais de BI.

No que tange ao volume, por exemplo, as ferramentas tradicionais de mineração de dados, em grande medida, tentam derivar valor a partir de grandes quantidades de *small data* – e não de *Big Data*. As ferramentas tradicionais de mineração de dados têm ignorado informações não

estruturadas e semiestruturadas, uma parte crescente do que hoje chamamos de *Big Data*. As pessoas e organizações têm usado os softwares de mineração de dados em conjuntos de dados mais ordenados do que os dados considerados não estruturados (DAVENPORT, 2014).

*Big Data* tem uma abordagem muito diferente em relação aos dados: o próprio processo de armazenamento e análise de dados é modificado em sua estrutura, de maneira que a tendência é que todos os dados gerados por uma organização são reunidos e interagem entre si. Isso permite que os administradores e analistas se preocupem sobre como usar os dados apenas em um segundo momento. Nesse sentido, as soluções de *Big Data* se provam ser mais escaláveis do que os bancos de dados tradicionais e armazéns de dados (OHLHORST, 2013).

Todas essas diferenças são técnicas. Do ponto de vista gerencial, as organizações que capitalizam sobre *Big Data* vão além de ambientes de análise de dados tradicionais de três maneiras fundamentais: elas prestam atenção ao fluxo de dados ao invés de se atentar aos estoques – o que se reflete nos processos de como elas derivam valor dos dados –, elas dependem de profissionais que têm habilidades além daquelas dos quais analistas de dados são dotados, e elas estão tornando a análise de dados um processo cada vez mais próximo ao núcleo de negócios e às funções operacionais e de produção – e cada vez mais distante do núcleo de tecnologia da informação (DAVENPORT e PATIL, 2012).

Assim, o *Big Data* surge como um modelo evolutivo da análise de dados tradicionais, suprimindo novas demandas de análise rápida dos dados advindos de diversas fontes e em maior quantidade. Empresas que já utilizam a análise de dados como ferramenta de competitividade têm vantagem no uso das novas ferramentas analíticas, a se destacar a melhoria nos processos produtivos, de gestão, de operação e de modelos de vendas e interação com os consumidores (NOVO e NEVES, 2013).

## **2.6. EVOLUÇÃO E DESENVOLVIMENTO DO *BIG DATA***

O termo “*Big Data*” é bem recente: o primeiro uso ocorreu em 2001, por Douglas Laney, então pesquisador do *Gartner Group Inc*. A popularização é ainda mais recente: o uso do Google Trends para investigar o surgimento do termo *Big Data* no histórico do sistema de busca do Google, como proposto por Choi & Varian (2009), nos revela um rápido crescimento em termos de interesse (número de busca) a partir de 2011 (BATTY, 2013).



**Figura 3:** Evolução do Interesse no termo *Big Data* ao longo do tempo (Fonte: autor)

Tendo em vista isso, nesta seção pretendo dissertar sobre o desenvolvimento conceitual e a evolução histórica do *Big Data*. Primeiro, vou introduzir o *framework* desenvolvido por Chen et alii (2012) que identifica a evolução e as aplicações e permitem a compreensão plena do conceito de *Big Data*. Em seguida, vou apresentar uma evolução histórica das tecnologias consideradas no *framework* criado por Chen et alii (2012). Por fim, será apresentado as características do momento atual da sociedade que favorecem o surgimento do *Big Data*.

### 2.6.1. Evolução Conceitual

*Business Intelligence & Analytics* (BI&A) e o campo relacionado à *Big Data analytics* têm apresentado importante crescimento tanto na comunidade acadêmica como na comunidade de negócios ao longo das últimas duas décadas. Tendo em vista isso, Chen et alii (2012) apresentaram um *framework* que identifica a evolução, as aplicações e o surgimento das diferentes áreas de pesquisa de BI&A. A seguir, apresento como os autores definiram e descreveram o que eles categorizaram como BI&A 1.0, BI&A 2.0 e BI&A 3.0.

#### **BI&A 1.0**

BI&A 1.0 tem suas raízes no campo de gestão de bases de dados. As tecnologias de análise de dados que integram as soluções de BI&A 1.0 em sua maioria dependem dos processos de extração, transformação e carga (ETL<sup>1</sup>) para conversão e integração de dados específicos da empresa. Os conjuntos de dados são, em sua maioria, estruturados, coletados pelas

<sup>1</sup> Do inglês “*Extraction, Transformation and Load*”

organizações por meio de vários sistemas legados, e usualmente armazenados em sistemas de gestão de banco de dados relacionais (RDBMS) comerciais. As técnicas de análise comumente utilizadas nestes sistemas, popularizado na década de 1990, estão fundamentadas principalmente em métodos estatísticos desenvolvidos na década de 1970 e as técnicas de mineração de dados desenvolvidos na década de 1980.

Sistemas de gestão de dados e *Data warehousing* são considerados os primeiros a integrar a categoria BI&A 1.0. Outros sistemas e ferramentas de consulta de banco de dados, processamento analítico online (OLAP<sup>2</sup>), e ferramentas de relatórios com base em gráficos simples e intuitivos para explorar características de dados importantes também são considerados como integrantes dessa categoria, bem como sistemas e ferramentas de gestão de performance do negócio (BPM<sup>3</sup>), que utilizam *scorecards* e *dashboards* para analisar e visualizar uma variedade de métricas de desempenho. Além dessas funções de relatórios de negócios bem estabelecidos, são adotadas técnicas de análises estatísticas e mineração de dados para análise de associação, segmentação e *clustering* dos dados, classificação e análise de regressão, detecção de anomalias e modelagem preditiva em várias aplicações de negócios.

Das 13 capacitações consideradas essenciais em plataformas de BI reportadas por Sallam *et al.* (2011) – conforme citação de Chen et alii (2012) –, as seguintes são consideradas como BI&A 1.0: relatórios e *dashboards*, queries ad hoc, sistemas baseados em busca de inteligência de mercados, OLAP, ferramentas de visualização interativa, scorecards, modelos preditivos e mineração de dados.

## **BI&A 2.0**

Desde o início de 2000, a Internet e a Web começaram a oferecer uma oportunidade única de coleta e investigação analítica de dados. Os sistemas de Web 1.0 baseados em HTTP caracterizados pelos motores de busca na Web, como o Google e o Yahoo, e empresas de comércio digital, como a Amazon e o Ebay, permitiram que as organizações apresentassem seus negócios online e interagissem diretamente com seus clientes. Além de suportar seus bancos de dados relacionais RDBMS tradicionais e seus conteúdos de negócios e produtos online, as novas plataformas levaram novas informações tais quais os logs do usuário, detalhes de IP continuamente recolhidos através de cookies e logs do servidor a se tornarem ferramentas para entender as necessidades dos clientes e identificar novos negócios e

---

<sup>2</sup> Do inglês “*Online Analytic Processing*”

<sup>3</sup> Do inglês “*Business Performance Management*”

oportunidades. Inteligência de rede, análise de rede e o conteúdo gerado pelos usuários coletados por meio da Web 2.0 – baseada em redes sociais e *crowdsourcing* – também se tornaram recursos valiosos e marcaram o início de uma nova era de pesquisa, o BI&A 2.0, na década de 2000, centrada na análise de texto e web para conteúdo não estruturados.

As aplicações da Web 2.0 foram desenvolvidas depois de 2004 e tem criado uma abundância de conteúdo gerado pelo usuário a partir de várias mídias sociais online, como fóruns, grupos online, blogs, redes sociais, sites de multimídia sociais (para fotos e vídeos), e até mesmo mundos virtuais e jogos online (O'REILLY, 2005 APUD CHEN, CHIANG e STOREY, 2012).

Diferentemente das tecnologias de BI&A 1.0, que já estão integradas em sistemas de tecnologias da informação de empresas comerciais, os futuros sistemas de BI&A 2.0 vão demandar a integração de técnicas maduras e escalonáveis de mineração de textos (como por exemplo, extração de informação, identificação de tópicos, mineração de opinião, resposta de consultas), mineração na web, análise de redes sociais e análises espaço-temporais integradas com sistemas de gestão de bases de dados relacionais das plataformas de BI&A 1.0.

### **BI&A 3.0**

Sistemas móveis de interface, visualização e interação computador-homem (HCI<sup>4</sup>) também são promissoras áreas de investigação e negócios. Embora o desenvolvimento da era do Web 3.0 (conexão móvel e baseada em sensores) seja certo, os sistemas de análise de conteúdo móvel, técnicas de localização e sistemas sensíveis ao contexto subjacentes para a coleta, processamento, análise e visualização de tais dados móveis fluídos e em larga escala ainda são desconhecidos.

Nenhum sistema comercial de BI&A 3.0 integrado é previsto para o futuro próximo. A maior parte da investigação acadêmica sobre BI móvel ainda está em um estágio embrionário.

Apesar de não ser incluído nas capacidades centrais plataforma de BI atuais, BI móvel foi incluído na análise do *Gartner BI Hype Cycle* como uma das novas tecnologias que tem o potencial de perturbar o mercado de BI de forma significativa (BITTERER, 2011 APUD CHEN, CHIANG e STOREY, 2012).

---

<sup>4</sup> Do inglês *Human-computer Interaction*

A figura a seguir sumariza as características chave das plataformas de BI&A 1.0, 2.0, 2.3.0 em relação as capacidades centrais das plataformas de BI da Gartner e do *Gartner BI Hype Cycle*.

	Principais Características	Gartner BI Platforms core capabilities	Gartner Hype Cycle
BI&A 1.0	Conteúdo estruturado, baseados em tecnologia DBMS: <ul style="list-style-type: none"> <li>• RDBMS e data warehousing</li> <li>• ETL e OLAP</li> <li>• Dashboards e scorecards</li> <li>• Mineração de dados e Análise estatística</li> </ul>	<ul style="list-style-type: none"> <li>• <i>Ad hoc</i> query e busca baseada em BI</li> <li>• Relatórios, dashboards e scorecards</li> <li>• OLAP</li> <li>• Visualização interativa</li> <li>• Modelagem preditiva e mineração de dados</li> </ul>	<ul style="list-style-type: none"> <li>• DBMS baseado em colunas</li> <li>• <i>In-memory</i> DBMS</li> <li>• Decisão em tempo real</li> <li>• Mineração de dados em bancada</li> </ul>
BI&A 2.0	Conteúdo desestruturado, baseado na WEB <ul style="list-style-type: none"> <li>• Extração e recuperação de informação</li> <li>• Mineração de opinião</li> <li>• Resposta à questionamentos</li> <li>• Análise da Web e Inteligência da Web</li> <li>• Análise de mídia social</li> <li>• Análise de rede social</li> <li>• Análise espaço-temporal</li> </ul>		<ul style="list-style-type: none"> <li>• Serviço de informação semântica</li> <li>• Resposta de linguagem natural aos questionamentos</li> <li>• Análise de texto e conteúdo</li> </ul>
BI&A 3.0	Conteúdo baseado em sensores e mobile <ul style="list-style-type: none"> <li>• Análise baseada na localização</li> <li>• Análise centrada no indivíduo</li> <li>• Análise de relevância do contexto</li> <li>• Visualização móvel e HCI</li> </ul>		<ul style="list-style-type: none"> <li>• BI baseado em tecnologia móvel</li> </ul>

**Figura 4:** Evolução do BI&A - Principais Características e Capacitações (Fonte: CHEN, CHIANG e STOREY, 2012)

### 2.6.2. Evolução Temporal da Análise de *Big Data*

Talvez o primeiro desafio de *Big Data* tenha sido o censo 1880 nos EUA, quando as informações relativas a aproximadamente 50 milhões de pessoas tiveram de ser recolhidos, classificados e comunicados. Esse desafio foi solucionado durante o censo de 1890, graças à introdução da primeira plataforma de *Big Data*: um dispositivo mecânico chamado a *Tabulating System Hollerith* que permitiu que a análise fosse feita em seis semanas, em vez de sete anos (OHLHORST, 2013).

O próximo salto gigantesco para o *Big Data* ocorreu com o *Manhattan Project* - o desenvolvimento dos Estados Unidos da bomba atômica durante a II Guerra Mundial -, que também foi o catalisador para a "Big Science" conceito que ganhou moeda com o programa espacial dos Estados Unidos no final dos anos 1950 e 1960 (OHLHORST, 2013).

Embora no mundo de hoje os censos de 1880-90 e o *Manhattan Project* não sejam mais considerado *Big Data* – pois computadores desktop de hoje podem analisar os dados sem grandes desafios –, eles evidenciam duas características do *Big Data*: primeiro, que o conceito envolve ter mais dados com os quais é possível lidar com a tecnologia da época; segundo, que



a atual onda de *Big Data* antes de fazer sucesso na comunidade de negócios tem suas raízes nas comunidades científica e médica (OHLHORST, 2013).

As raízes do envolvimento da comunidade de negócios com *Big Data* datam das décadas de 1970 e 1980. Em 1970, grandes empresas de bens de consumo foram as pioneiras no cruzamento de dados de auditores do mercado externo, como a Nielsen, com dados internos para analisar, planejar, executar e medir as suas estratégias de marketing e publicidades, incluindo gastos com promoção e decisões de preço (SCHMARZO, 2013).

Na década de 1980, as tecnologias como scanner de dados de ponto-de-venda (POS<sup>5</sup>) provisionaram varejistas com uma visão única sobre as vendas de produtos, padrões de compra dos clientes e tendências do mercado global, graças a combinação dos sistemas POS do varejo com o uso de códigos de barras. O volume de dados no varejo saltou de *megabytes* para *gigabytes* e logo a *terabytes*, o que exigiu uma nova geração de capacidade de processamento de dados, representada por novas arquiteturas cliente-servidor e plataformas de dados como Teradata, Sybase IQ e Informix. Era o nascimento da indústria de software de *Business Intelligence* (BI), a partir do projeto de “apoio à decisão” da Procter & Gamble no final de 1980 (SCHMARZO, 2013).

Em seguida, aplicações e sistemas para usuários da ponta da cadeia, tais quais sistemas MRP<sup>6</sup>, começaram a ser desenvolvidos com base no modelo de dados relacionais. Em um nível alto, esses sistemas tinham como objetivo automatizar processos padrões de negócios. Para isto, enormes bancos de dados foram desenvolvidos. Tais como seus antecessores, sistemas ERP<sup>7</sup> e CRM destacavam-se apenas na manipulação de dados estruturados, com a distinção que agora executavam também outras funções essenciais de negócios e forneciam relatórios padrão (SIMON, 2013).

Ao longo das últimas décadas – 1990 e 2000 –, cada vez mais organizações desenvolveram sistemas construídos sobre este modelo de dados relacionais. Investiu-se cada vez mais tempo, dinheiro e esforço à gestão de dados de natureza interna, isto é, aos dados gerados pelas próprias organizações (SIMON, 2013).

Ainda no final da década de 1990, registros da web permitiram revelar novos insights sobre vendas de produtos e comportamentos de compra dos clientes e, novamente, as empresas tiveram que mudar seus modelos de negócio. Mas a adaptação dos sistemas foi relativamente

---

<sup>5</sup> Do inglês “*Point-of-Sale*”.

<sup>6</sup> Do inglês, “*Manufacturing Resource Planning*”.

<sup>7</sup> Do inglês, “*Enterprise Resource Planning*”.

natural, tendo em vista o surgimento dos sistemas de processamento analítico online (OLAP<sup>8</sup>), utilizados para análise de tabelas de dados multidimensionais online (SCHMARZO, 2013).

Hoje, estamos no meio de mais uma revolução de negócios orientada a dados: novas fontes de dados, tais como mídias sociais, tecnologias móveis e sensores têm o potencial de mudar o processo de criação de valor de uma organização. Os volumes massivos, a diversidade e complexidade dos dados estão forçando as capacidades tecnológicas (SCHMARZO, 2013).

Essa nova revolução ganhou força por volta de 2005, quando a Web 2.0 – também conhecida por social web – tornou-se realidade. Como resultado direto, o volume, a variedade e a velocidade de dados cresceram de forma exponencial, especialmente de dados voltados para o consumidor, que é, em sua maior parte, externo à empresa. Novos sites como Google, Facebook e outros sites e plataformas sociais levaram à proliferação de dados não estruturados, semiestruturados e metadados (SIMON, 2013).

Em resumo, o termo *Business Analytics* foi introduzido para representar os componentes analíticos principais do *Business Intelligence* (DAVENPORT, 2006). Mais recentemente, *Big Data* e *Big Data analytics* têm sido utilizados para descrever conjuntos de dados e técnicas analíticas que são tão grandes e tão complexas que requerem armazenamento, gerenciamento, análise e tecnologias de visualização únicas e avançadas (CHEN, CHIANG e STOREY, 2012). A tabela a seguir, extraída de Davenport (2014), resume a evolução histórica da análise de dados. No anexo III, há uma figura criada pelo próprio autor que indica a evolução conceitual e temporal de maneira conjunta.

---

<sup>8</sup> Do inglês, “*Online Analytical Processing*”.

<b>Termo</b>	<b>Época</b>	<b>Significado específico</b>
Decision support	1970-1985	Uso da análise de dados para suporte à tomada de decisões.
Executive Support	1980-1990	Foco na análise de dados para decisões tomadas pelos executivos seniors.
Online analytical processing (OLAP)	1990-2000	Software para a análise de tabelas de dados multidimensionais.
Business intelligence	1989-2005	Conjunto de ferramentas de suporte a decisões baseadas em dados, com ênfase no report.
Analytics	2005-2010	Foco em análises estatísticas e matemáticas para a tomada de decisões.
<i>Big Data</i>	2010-present	Foco em dados muito grandes, desestruturados e contínuos.

**Tabela 2:** Evolução Temporal da Terminologia de Análise de Dados (Fonte: DAVENPORT, 2014).

## 2.7. BENEFÍCIOS E OPORTUNIDADES<sup>9</sup>

Google, eBay e LinkedIn foram algumas das primeiras organizações a utilizar o *Big Data*, implementando e desenvolvendo provas de conceito e projetos de pequena escala para aprender se seus modelos de análise de dados poderiam ser melhores por meio do uso de novas fontes de informação (DAVENPORT, 2014).

As empresas que dominam a gestão orientada para dados podem obter recompensas significativas e se diferenciar de seus concorrentes (CHEN, CHIANG e STOREY, 2012).

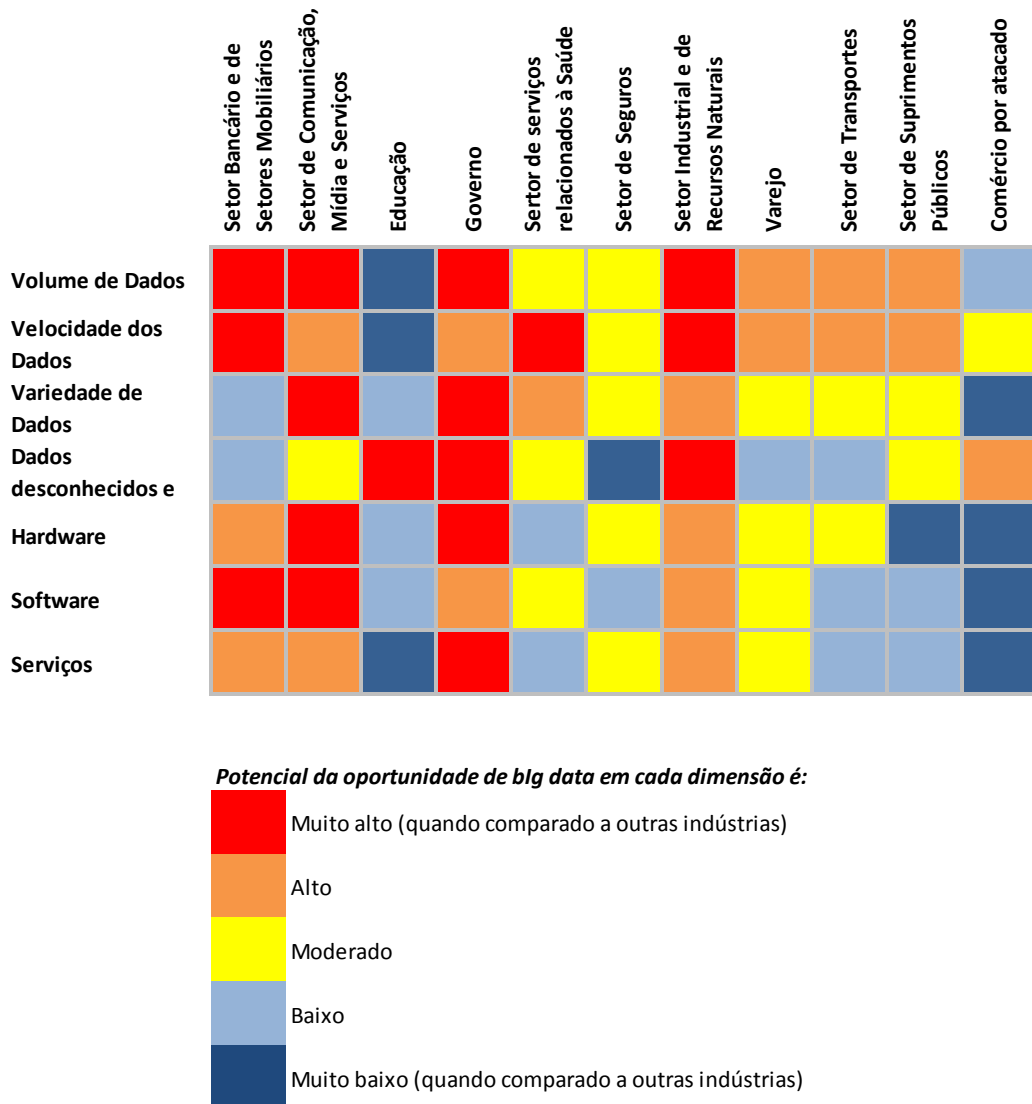
<sup>9</sup> Vale destacar que não há clareza na literatura específica sobre o que diferencia um benefício de uma oportunidade, havendo por vezes utilização de um dos termos quando o mais adequado seria o outro. Pelo dicionário, a diferença entre os termos resume-se a uma questão temporal: benefícios são consequências positivas de oportunidades, sendo que essas podem ou não ser aproveitadas. No presente trabalho serão considerados benefícios toda a vantagem, proveito ou provento obtido por uma organização a partir da análise dos dados estruturados e não estruturados, internos e externos, em quaisquer umas de suas esferas. Por oportunidades entende-se benefícios potenciais e estas residem no estudo em questão nos projetos totalmente fundamentados em análise de dados.

O presente capítulo objetiva reunir alguns dos benefícios e oportunidades que podem levar as organizações a adotar o *Big Data*. Principalmente porque, embora sejam significativos, as oportunidades e os benefícios podem ser difíceis de serem identificados antes de um projeto de análise de dados ser iniciado (NOVO e NEVES, 2013; SILVA e CAMPOS, 2014).

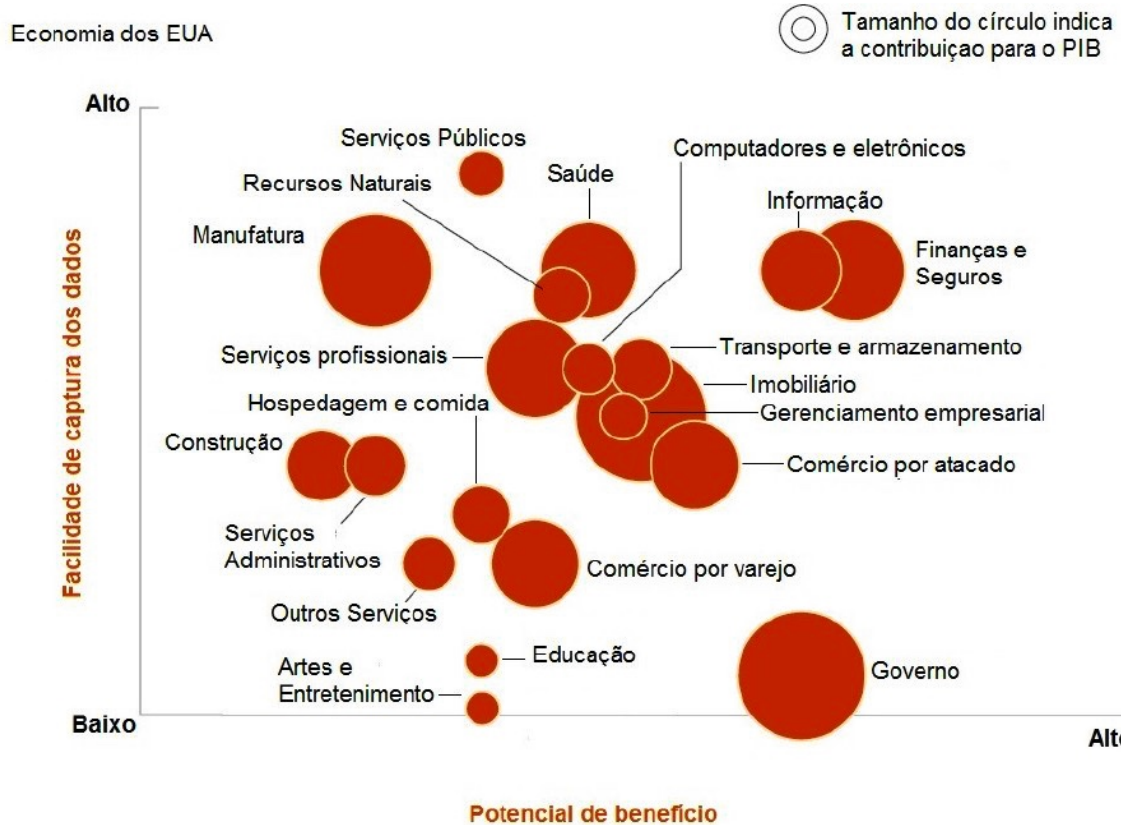
O capítulo está organizado da seguinte maneira: primeiro, será exposto que os benefícios e oportunidades advindos do *Big Data* podem ser encontrados em qualquer indústria e setor econômicos, apresentando alguns exemplos; A seguir, serão apresentados alguns benefícios e oportunidades gerenciais da tecnologia – isto é, que podem ser aplicados e usufruídos em qualquer indústria, organização ou departamento de uma organização; depois, serão apresentadas algumas oportunidade nos mais diversos setores e departamentos de uma organização.

### **2.7.1. Benefícios e oportunidades setoriais**

*Big Data* tem demonstrado a capacidade de melhorar as previsões, economizar custos, aumentar a eficiência e melhorar a tomada de decisões em áreas tão díspares como controle de tráfego, previsão do tempo, prevenção de desastres, finanças, controle de fraudes, transação comercial e educação, dentre outros (YAN, 2013). Os benefícios não são disponíveis apenas para as empresas privadas. Os governos também esperam que *Big Data* contribua para melhorar a sua capacidade para servir os cidadãos e enfrentar os grandes desafios públicos que envolvem a economia, saúde, criação de emprego, desastres naturais e terrorismo (KIM, TRIMI e CHUNG, 2014). Em suma, os benefícios e oportunidades oriundos do *Big Data* podem ser encontrados ao longo de toda a economia – em diversas indústrias e setores econômicos –, conforme destacam as figuras 5 e 6 a seguir.



**Figura 5:** Mapa de Oportunidades do *Big Data* por Indústria (Fonte: YAN, 2013)



**Figura 6:** Facilidade de captura dos dados versus Potencial de benefício, por indústria (Fonte: NOVO & NEVES, 2013)

A seguir são apresentadas algumas oportunidades, projetos e benefícios de alguns dos setores da economia: (1) governo, (2) ciência e tecnologia, (3) saúde e bem-estar, e (4) segurança pública e privada.

### GOVERNO

O advento da Web 2.0 tem colaborado para a reinvenção dos governos. A campanha presidencial de 2008 nos EUA e as campanhas eleitorais para o senado seguintes deram os primeiros sinais de sucesso para campanhas e participações políticas on-line. Apelidada de "política 2.0", essa nova era é representada por políticos que usam plataformas web para discussões de políticas bem-sucedidas, publicidade de campanha, mobilização de eleitores, anúncios de eventos e doações on-line. E nesse processo em que o governo e os processos políticos tornam-se mais transparentes, participativos, online e ricos em multimídia, há uma grande oportunidade para a adoção do *Big Data* no governo. Técnicas selecionadas de mineração de opinião, análise de rede social e análise de dados em mídias sociais podem ser

usados para apoiar a participação política on-line, a democracia eletrônica, a prestação de serviços de governo eletrônico e transparência dos processos e prestação de contas (CHEN, 2009 APUD CHEN, CHIANG e STOREY, 2012).

Serviços governamentais ou do setor público são conhecidos pela criação de grandes volumes de dados. *Big Data* proporciona, assim, uma abundância de oportunidades para os governos. De acordo com grupo de estudos de livre mercado sobre política de comércio do Reino Unido, o governo do Reino Unido poderia economizar até £33 bilhões por ano usando *Big Data* para se tornar mais eficaz. A consultoria McKinsey investigou que o valor anual potencial de benefícios para setor público europeu é de 250 bilhões de euros graças ao *Big Data* (VAN RIJMENAM, 2014), embora o impacto atual no setor público ainda seja relativamente pequeno frente a esse potencial, de acordo com o grupo de TI MeriTalk (SHUEH, 2014).

De acordo com Van Rijmenam (2014), os 4 principais benefícios que o governo pode extrair do *Big Data* são: (1) melhora da transparência e da tomada de decisão concomitante à redução de custos, (2) personalização da experiência dos cidadãos, (3) redução da fraude nos impostos e na seguridade social, e (4) melhora na segurança pública e na saúde. Embora alguns tópicos sejam similares, a StateTech Magazine (2013) fez uma lista ainda mais extensa de benefícios do *Big Data* para o governo: (1) melhora na tomada de decisão, (2) melhora de resultados dos projetos governamentais, (3) identificação e redução das ineficiências, (4) eliminação do desperdício, fraude e abuso, (5) aumento da produtividade, (6) diminuição do custo total de propriedade e melhora do retorno sobre os investimentos, (7) melhora da transparência e dos serviços prestados, e (8) redução da criminalidade e das ameaças à segurança pública.

Reconhecendo a importância de *Big Data*, a Casa Branca formou o Grupo Diretor Sênior de *Big Data* em 2010 e, dois anos mais tarde, anunciou a Iniciativa de Pesquisa e Desenvolvimento em *Big Data* para “ajudar a acelerar o ritmo de descoberta em ciência e engenharia, reforçar a segurança nacional, e transformar o ensino e a aprendizagem, melhorando nossa capacidade de extrair conhecimento e insights de grande e coleções complexas de dados digitais”. A iniciativa apresenta 11 programas de seis agências e financiamento de US\$ 200 milhões para apoiar atividades de *Big Data* em todo o Poder Executivo (THE U.S. OFFICE OF PERSONNEL MANAGEMENT, 2012).

Outra iniciativa norte-americana que vale ser destacada é o Data.gov, um projeto do governo federal de acesso livre a dados públicos, em que podem ser encontrados dados americanos federais, estaduais e locais fornecidos por centenas de organizações, além de ferramentas e

recursos para conduzir pesquisas, construir aplicativos e ferramentas de visualização e design da informação. Ao utilizar tecnologias de *Big Data*, o Data.gov também pode contribuir para a oferta de melhores serviços ao público em geral (YAN, 2013).

O EUA não é o único país cujo governo usufrui dos benefícios do *Big Data*. O Brasil também já teve experiências com o *Big Data*, embora em uma escala muito menor quando comparada à norte-americana. Um estudo de caso realizado para o Fundo Nacional de Desenvolvimento da Educação (FNDE) a respeito da compra de antibióticos para Hospitais Universitários Federais, por exemplo, resultou em economia de R\$10,6 milhões (BLUMENSCHHEIN, 2014).

Ainda de acordo com Blumenschein (2014), o uso de *Big Data* também pode ser usado para ampliar o escopo das análises realizadas na elaboração de estratégias de aquisições governamentais, tanto para caracterizar a demanda como o produto, de modo a garantir uma melhor compreensão do perfil de estrutura de custos do setor. Além disso, o *Big Data* pode ser usado no mapeamento das empresas que podem ofertar os produtos, como para estabelecer o preço de referência, e, por conseguinte, contribuir para satisfazer as necessidades da administração pública ao mesmo tempo em que otimiza o uso dos recursos públicos.

Além do uso para pesquisas, o Brasil também tem uma iniciativa com *Big Data* para visualização de dados à semelhança do Data.gov. O Dataviva é um instrumento de visualização de dados que pode ser acessado de forma gratuita por qualquer pessoa interessada em “informações sobre as atividades econômicas, ocupações e exportações de cada um dos 5.561 municípios brasileiros nos últimos dez anos”. O Dataviva foi desenvolvido pelo Escritório de Prioridades Estratégicas do Governo de Minas Gerais, em parceria com pesquisadores do MIT Media Lab (BARRENCE, GOMES e FREITAS, 2014).

Sem pretensão de esgotar os exemplos, o Brasil ainda tem uma iniciativa no poder público municipal da cidade do Rio de Janeiro – o Pensa, que tem o intuito de analisar o grande volume de dados gerados pela sociedade e assim desenvolver os serviços prestados conforme as necessidades de seus cidadãos. Essa iniciativa é o objeto de estudo deste trabalho e será mais bem caracterizada posteriormente.

Outras iniciativas ao redor do mundo são o *UK Horizon Scanning Center* (HSC), estabelecido em 2004 para melhorar a habilidade do governo britânico em lidar com desafios multidisciplinares interdepartamentais e o IBM DOME, projeto colaborativo da IBM com o governo de 20 países, incluindo o Reino Unido, a Suíça e a Holanda, que objetiva desenvolver o sistema de um supercomputador capaz de analisar até um exabyte/dia de dados oriundos do



projeto de *Big Data* de astronomia chamado *Square Kilometer Array* (SKA) (KIM, TRIMI e CHUNG, 2014). Kim et alii (2014) descrevem iniciativas governamentais em *Big Data* em outros países ao redor do mundo, tais como Japão, Austrália, Coréia e Singapura, conforme pode ser visto no anexo IV. Estas iniciativas são apenas ilustrativas, pois se multiplicam rapidamente.

### CIÊNCIA E TECNOLOGIA

Muitas áreas da Ciência e Tecnologia (C&T) estão colhendo os benefícios de sensores de alto rendimento e instrumentos, de astrofísica e oceanografia à genética e investigação ambiental (CHEN, CHIANG e STOREY, 2012).

Para facilitar o compartilhamento de informações e de dados de análise, a *National Science Foundation* (NSF) determinou recentemente que cada projeto deve apresentar um plano de gerenciamento de dados, com intento de promover os meios científicos e tecnológicos de gestão, analisar, visualizar e extrair informações úteis de grandes conjuntos de dados, diversos, distribuídos e heterogêneos, de modo a acelerar o progresso da descoberta científica e da inovação, incentivar o desenvolvimento de novas ferramentas de análise de dados e algoritmos, facilitar a infraestrutura de dados escaláveis, acessíveis e sustentáveis, aumentar a compreensão dos processos e interações humanas e sociais, e promover o crescimento econômico e a melhoria da saúde e qualidade de vida (CHEN, CHIANG e STOREY, 2012).

Outro projeto que merece destaque é o ENCODE, que gerou uma incrível quantidade de informação por causa de sua escala: mais de 1.600 experimentos sobre 147 tipos de células, 235 tipos de anticorpos e outros protocolos de ensaio. O projeto em si contou com a contribuição de cerca de 450 autores, atuantes em mais de 30 instituições e tem como objetivo fazer da ciência um consórcio: os pesquisadores devem se concentrar em criar o melhor conjunto de dados que podem - e por "melhor conjunto de dados", não se resume a apenas dados brutos, mas também análises intermediárias que permitem aos cientistas escolher o nível de detalhe em qual pretendem iniciar. Talvez os dados sejam utilizados, talvez não. O importante é o recurso para a comunidade, não o sucesso individual (BIRNEY, 2012). De acordo com Birney (2012), pode-se observar nesse projeto benefícios como a maior produção intelectual, a possibilidade de cooperação de pesquisadores em todo o mundo, o acesso a dados de pesquisas realizadas em outros lugares do mundo para repetir um experimento sem necessariamente ter que formulá-lo desde o início, entre outros benefícios.

Saindo do cotidiano e observando o ambiente de ciência e tecnologia, existem muitos outros exemplos de onde é possível identificar outros benefícios, como, por exemplo, a viabilidade de novas descobertas e novas pesquisas. Os projetos de pesquisa de perfuração de petróleo em águas profundas, incluindo o pré-sal (SANTOS, 2014), além de projetos de pesquisa em astronomia, são exemplos de outros projetos que estão impulsionando o uso da abordagem de *Big Data* e de como a ciência podem se beneficiar da tecnologia (XEXÉO, 2013; PORTO, 2013).

De fato, os artigos com maior número de citações sobre o *Big Data* são da área de pesquisa médica, com destaque para pesquisas biológicas, bem como os autores de maior destaque quando o tema é *Big Data* também publicaram trabalhos na área médica, com possíveis aplicações em sistemas biológicos, epidemiologia, psicologia e doenças cardíacas (SILVA e CAMPOS, 2014).

### SAÚDE

As principais fontes de dados de dados de saúde são dados genéticos oriundos de iniciativas de *Big Data* na área (relacionados ao mapeamento genético, à expressão gênica, e ao sequenciamento de dados) e dados de prestadores de serviços de saúde (registros de saúde eletrônicos, registros seguros, farmácia da prescrição, feedback paciente e respostas) (MILLER, 2013). A partir dessas fontes é possível gerar benefícios tais como o melhor conhecimento do paciente (estilo de vida e como a doença impactará a vida do paciente) – que permite a criação de planos de saúde mais customizados e melhora do relacionamento com os pacientes do setor –, abordagens sustentáveis que continuamente garantirão valor por meio da redução de custos e melhora nos processos de inovação, e aumento e melhora das bases de dados e conhecimento de saúde (WACTLAR, PAVEL e BARKIS, 2011; CHEN, 2011; GROVES, KAYYALI, *et al.*, 2013)

Como exemplos de como *Big Data* pode alterar a relação com o consumidor podem ser citados os sites *Daily Strength* e *PatientsLikeMe*, que proporcionam oportunidades de pesquisa únicas no apoio à tomada de decisão na Saúde e na capacitação dos doentes, especialmente para doenças crônicas, como diabetes, mal de Parkinson, Alzheimer e câncer (MILLER, 2013).

Outros benefícios potenciais incluem a detecção de doenças em estágios iniciais, quando eles podem ser tratados mais facilmente e de forma eficaz; gestão da saúde individual e populacional específica e detecção da fraude de cuidados de saúde de forma mais rápida e

eficiente (RAGHUPATHI e RAGHUPATHI, 2014). Dentre os benefícios mais econômico-financeiros podem ser citados o aumento da eficiência operacional, melhora na captura de informação estratégica, identificação de novos mercados, ida mais rápida ao mercado de novos medicamentos, melhor atendimento das regulações implementadas pelo governo, entre outras (CHEN e ZHANG, 2014).

A níveis regionais, o relatório CERIC identificou como benefício a melhoria da vigilância da saúde pública e da resposta à incidentes e epidemias, melhor segmentação de serviços urgentes à saúde pública, melhoria da segurança dos procedimentos e diminuição dos erros médicos, entre outros benefícios. Um estudo recente da McKinsey afirma que, com *Big Data analytics*, estima-se que o custo anual total do tratamento para pacientes com risco de doença cardíaca coronariana poderia ser reduzido em mais de \$30 bilhões, tendo apenas algumas intervenções simples, como o uso de aspirina combinada com rastreio precoce colesterol e cessação do tabagismo (GROVES, KAYYALI, *et al.*, 2013).

## SEGURANÇA

Desde os trágicos acontecimentos de 11 de setembro de 2001, a área de segurança ganhou muita atenção e as questões de segurança são uma grande preocupação para a maioria das organizações. Agências de Inteligência, Segurança e Segurança Pública estão reunindo grandes quantidades de dados de múltiplas fontes para desenvolver soluções contra as ameaças de segurança, inclusive cibernética (CHEN, CHIANG e STOREY, 2012). O volume e a variedade de dados em questão são tão grandes que só mesmo *Big Data* para permitir que a análise de todos os dados seja factível. Para entender à dimensão do volume e da variedade, basta considerar que o *Patriot Act* dos EUA autoriza qualquer órgão do governo a investigar ou se envolver em atividades de inteligência relacionados com o terrorismo internacional, além de conferir poder para emissão das chamadas "cartas de segurança nacional" para obrigar a divulgação de um relatório sobre todas as informações no arquivo de um consumidor, se a agência certificar por escrito que o relatório é necessário para a investigação, atividade, ou análise (DEMPSEY e FLINT, 2004).

Reconhecendo os desafios apresentados pelo volume e complexidade dos projetos de *Big Data* relacionados com a defesa, a Agência americana de Projeto de Pesquisa Avançada em Defesa (DARPA<sup>10</sup>) iniciou em 2012 um programa em conjunto com o Departamento de Defesa, o XDATA, para ajudar a desenvolver técnicas computacionais e ferramentas de

---

<sup>10</sup> Do inglês *US Defense Advanced Research Project Agency*.

software para processar e analisar a vasta quantidade de dados orientados para atividades de defesa. O XDATA visa abordar a necessidade de algoritmos escaláveis para processamento e visualização de dados imperfeitos e incompletos. O programa envolve matemática aplicada, informática, e comunidades de visualização de dados para desenvolver soluções de usabilidade para combatentes de guerra (CHEN, CHIANG e STOREY, 2012).

Os benefícios do *Big Data* para as questões de segurança vão além da capacidade de processar grandes conjuntos de dados das fontes mais variadas e incluem, por exemplo, a capacidade de identificação de padrões e a possibilidade de análise em tempo real. *Big Data* pode expor padrões de comportamento ocultos das pessoas e até mesmo lançar luz sobre suas intenções. Esta informação é útil para as agências governamentais, bem como as empresas privadas a apoiar a tomada de decisão em áreas da segurança. É particularmente interessante para áreas aplicadas da consciência situacional e as abordagens preventivas necessárias para a descoberta quase em tempo real (MICHAEL e MILLER, 2013).

*Big Data* também pode contribuir para a Segurança Nacional alterando os processos de vigilância de três maneiras principais. Em primeiro lugar, dado que *Big Data* envolve o uso ampliado de algoritmos para análise, há uma crescente dependência de software para vigilância e uma dependência concomitante do relacionamento “homem-algoritmo” que molda as maneiras que seres humanos são tratados por sistemas de vigilância. Em segundo lugar, com as práticas de *Big Data*, as operações de vigilância concentram-se cada vez mais sobre o futuro do que no presente e no passado. No contexto da governação neoliberal, essa antecipação é susceptível de colocar mais peso sobre a vigilância da gestão das consequências em vez de pesquisa sobre a compreensão das causas dos problemas sociais, tais como o crime e a desordem. A terceira área é a adaptação, partindo do pressuposto que um processo ou método de análise de dados bem-sucedido ou busca de dados podem ser transferidos com sucesso e pouco risco de um campo para outro (LYON, 2014).

Nesse contexto, para maximizar os benefícios que o *Big Data* pode gerar para a segurança pública, é necessário que as organizações passem a observar o que não é usual, adote maiores conjuntos de dados, integre o uso de análise de dados à rápidas respostas, deixe de olhar apenas para as ameaças conhecidas, considere soluções de segurança baseadas em propósitos, não espere que o análise de dados substitua à experiência dos especialistas e inicie projetos de desenvolvimento de Tecnologias de Análise de Segurança (ROBB, 2014).

### **2.7.2. Benefícios e Oportunidades de Gestão**

Três classes de valor são mais usuais em *Big Data*: redução de custos, melhorias de decisão e melhoria de produtos e serviços (DAVENPORT, 2014). A McKinsey Global Institute (2011) e Tankard (2012) acrescentam a essa lista de benefícios a criação de transparência nas atividades organizacionais que podem ser usados para aumentar a eficiência e a melhora nos processos de inovação. Embora intimamente associado aos benefícios já listados, Silva & Campos (2014) acrescenta ainda a melhora nas previsões de demanda de mercado e a capacidade de desenvolver vantagens competitivas mais sustentáveis.

O *Big Data Executive Survey* (2013) cita outros benefícios, mais a nível operacional e intimamente relacionados com a definição de *Big Data*, como, por exemplo, a integração de dados de diversas fontes, análise de grandes volumes de dados, melhora das capacitações analíticas gerais, análise de informação em tempo real e redução de custos da análise e processamento de dados.

No presente tópico, irei explorar os seguintes benefícios: (1) redução dos custos e aumento das receitas; (2) aumento da eficiência operacional; (3) melhora na tomada de decisão; (4) melhora de produtos e serviços já existentes; e (5) a melhora nos processos de inovação e o desenvolvimento de novos produtos e mercados. Essa lista de benefícios gerais não objetiva ser exaustiva. A tabela a seguir indica quais autores escrevem sobre esses benefícios.

Autores	Benefícios			
	Aumento da eficiência operacional e melhor desempenho financeiro	Melhora na tomada de decisão	Inovação incremental em produtos e serviços já existentes, e aumento da lealdade dos clientes	Inovação disruptiva de produtos, serviços e modelos de negócios
Bloem <i>et. al.</i> (2012)			x	
Brown, Chui e Manyika (2011)		x	x	
Davenport (2014)	x	x	x	x
Gallant (2011)		x	x	
Leeflang, Verhoeff <i>et. al.</i> (2014)	x		x	
McAfee e Brynjolfsson (2012)	x		x	x
McKinsey Global Inst. (2011)	x	x		x
Minelli <i>et. al.</i> (2013)	x	x		
Novo e Neves (2013)	x		x	x
Ohlhorst (2013)			x	
Silva & Campos (2013)	x			x
Tankard (2012)	x	x		x
Ularu, Puican <i>et. al.</i> (2012)			x	x
Yan (2013)	x			

**Tabela 3:** Benefícios Gerais citados por autores (**Fonte:** autor)

### AUMENTO DA EFICIÊNCIA OPERACIONAL E MELHOR DESEMPENHO

De acordo com o McKinsey Global Institute, um varejista que aproveita adequadamente *Big Data* tem o potencial de aumentar suas margens operacionais em mais de 60% e ganhar *market share* dos seus concorrentes por aproveitar dados detalhados de clientes (MCKINSEY GLOBAL INSTITUTE, 2011). Brynjolfsson (2012) descreve um estudo feito no MIT Center for Digital Business objetivando comprovar a hipótese de melhor desempenho das empresas movidas a dados. Para isto foram aplicadas entrevistas estruturadas com executivos de 330 empresas norte-americanas de capital aberto sobre suas práticas de gestão organizacional e TI (Tecnologia da Informação), agregando o resultado destes aos dados de desempenho disponíveis em balanços anuais e fontes independentes. Foi identificado que, dentre as empresas que mais faziam uso de dados nas tomadas de decisão, aproximadamente um terço das empresas tinha em média 5% mais produtividade e uma rentabilidade 6% maior do que as demais (NOVO e NEVES, 2013; SILVA e CAMPOS, 2014).

O aumento da eficiência operacional pode ocasionar uma redução dos custos, um aumento das receitas ou ambos aos mesmo tempo. De fato, o potencial da economia de custos devido ao *Big Data* é potencialmente alto. Ele está associado à mudança de plataforma para armazenamento e processamento de dados, que passa de uma grande e potente plataforma de dados estruturados para o uso de múltiplos servidores. Mesmo a armazenagem e o processamento de dados estruturados são mais baratos quando processados por meio de tecnologias de *Big Data*. Para comparação os custos de armazenagem de um *terabyte* por ano para uma companhia são de \$37 mil dólares em base de dados relacionais tradicionais, \$5 mil para *data appliance*, e apenas \$2 mil para *Hadoop cluster* (DAVENPORT, 2014).

As reduções de custo associadas ao aumento da eficiência operacional são oriundas de projetos de *Big Data* são potencialmente bem substanciais e podem contribuir para uma grande economia de custos e grande receita (DAVENPORT, 2014). Estas reduções de custo não passam apenas pelo custo de realização da análise de dados, mas pode ser consequência do investimento nas tecnologias de *Big Data*. O serviço de receita interna norte-americano<sup>11</sup>, por exemplo, usou análise de dados por meio de *Big Data* para corrigir erros no início do processo de arquivamento de impostos para 2012, economizando US\$ 100 milhões em créditos errôneos (FRANZEN, 2012 APUD YAN, 2013). No setor de transporte, a *US Xpress* economizou mais de US\$ 6 milhões por ano devido à diminuição do tempo ocioso de seus caminhões, garantindo que eles fossem mantidos em movimento (BIG DATA STARTUPS, 2012).

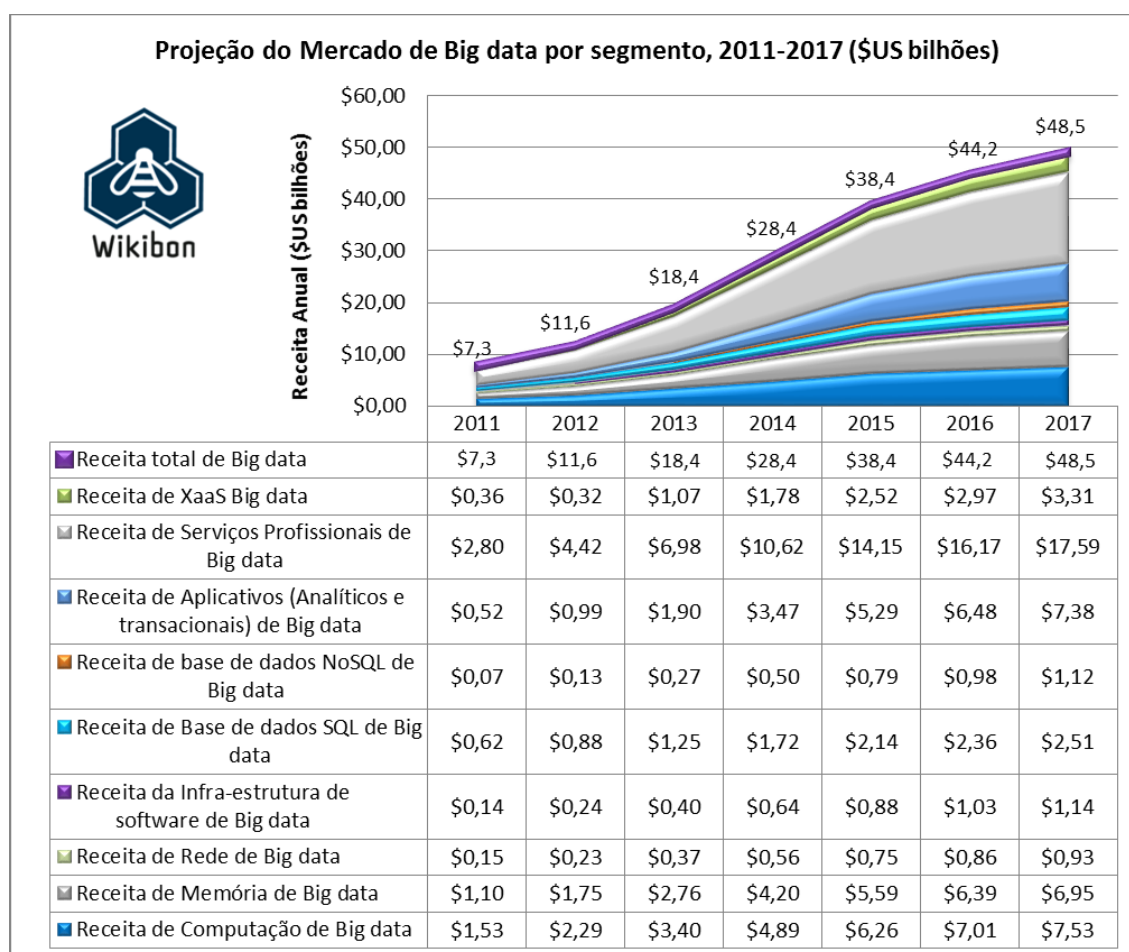
Uma grande empresa de petróleo, por exemplo, conseguiu reduzir os custos com pessoal em quase 25% e ao mesmo tempo aumentar a produtividade em 5% graças à adoção da tecnologia. Este ganho de eficiência operacional foi possível utilizando sensores ao longo da linha de produção, enviando dados que eram analisados em tempo real, permitindo ajustes imediatos na linha, evitando assim intervenções manuais, que são, em geral, demoradas, caras e perigosas (NOVO e NEVES, 2013).

Por outro lado, *Big Data* pode entregar maiores receitas provenientes de maior custo-eficiência para as empresas (MINELLI, CHAMBERS e DHIRAJ, 2013). Na indústria de TI, a IBM espera que sua receita de análise e tecnologias de *Big Data* chegue a US\$ 20 bilhões até 2015 (DOW JONES NEWSWIRE, 2012 APUD YAN, 2013). A Holiday Extras, uma companhia britânica de agendamento de hotéis e estacionamento em aeroportos alcançou receita anual de £ 200 milhões graças ao *Big Data* (BLOOMBERG, 2012). As receitas

---

<sup>11</sup> Em inglês, *U.S. Internal Revenue Service*

globais de mercado são projetadas para ser 48,5 bilhões dólares em 2017 (FLOYER, KELLY, *et al.*, 2013), como pode ser visto na figura abaixo.



**Figura 7:** Projeção de Mercado do *Big Data* por segmento em milhões de US\$, 2011-2017 (Fonte: © Wikibon *Big Data Model 2011-2017*)

Leeflang, Verhoeff et alii (2014) afirmaram que as empresas que analisam dados sistematicamente superam o desempenho de suas concorrentes, citando exemplos para embasar sua afirmação: Amazon.com (taxa de crescimento anual – AGR<sup>12</sup> – 2000-2010: 56,5%), O2 (AGR 29,5%), CapitalOne (AGR 16,6%), Tesco (AGR 11,7%), e progressiva (AGR 6%).

Destes exemplos, é válido aprofundar-se um pouco mais sobre o caso de negócios da varejista britânica Tesco. A Tesco construiu uma cultura de tomada de decisão baseada nos dados de seus clientes em todos os níveis da empresa para se tornar uma das maiores varejistas do mundo. A transformação sistemática de dados de cartões de fidelidade em *insights* – inclusive

<sup>12</sup> Do inglês, “*Annual Growth Rate*”.



para as decisões de negócio –, alimentou a ascensão da Tesco para a posição de maior varejista do Reino Unido. Para alcançar seus objetivos, a Tesco criou um poderoso mecanismo de coleta de dados por meio da combinação dos dados obtidos a partir de cartões de fidelidade, scanners, sites e pesquisas de mercado, para entender melhor os clientes. Finalmente, eles desenvolveram segmentações de clientes multidimensionais de acordo com necessidades e comportamentos semelhantes e comportamentos, para desenvolver conhecimentos específicos do segmento (LEEFLANG, VERHOEF, *et al.*, 2014).

### **MELHORA DA TOMADA DE DECISÃO**

O primeiro propósito por trás de uma análise de dados tradicional era o suporte à tomada de decisões interna dos negócios. Que proposta devemos apresentar para o cliente? Quais clientes tem maior tendência de deixar de ser clientes rapidamente? Quanto de estoque devemos manter? Como devemos precificar nossos produtos? Estes tipos de decisões são facilitadas com *Big Data* devido ao surgimento dos novos conjuntos de dados menos estruturados que podem ser aplicados a elas (DAVENPORT, 2014).

Com casos de uso em ambos os setores público e privado, os líderes do governo, da indústria e de negócios começaram a ver que as tecnologias avançadas em *Big Data* lhes permitiriam adquirir *insights* mais profundos e mais valiosos e tomar decisões com mais informação e em um *timing* melhor do que antes (MANYIKA, CHUI, *et al.*, 2011; ZIKOPOULOS, DEROOS, *et al.*, 2012; THE WHITE HOUSE, 2012).

Uma das áreas que pode se beneficiar do *Big Data* para melhorar a tomada de decisão é a área de Recursos Humanos (RH), sobretudo para a contratação de funcionários. No relatório, Previsões para 2013 da Consultoria Deloitte, o autor Bersin previu que a análise de dados, o *Big Data* e o planejamento da força de trabalho se tornariam o foco de RH para grandes equipes, ao permitir a melhor tomada de decisões. O autor escreve:

"Análise de *Big Data* merecem o status buzzword que conseguiram. Em 2012, os CEOs podem forçar RH para ficar atento às análises de dados e empresas podem impulsionar suas equipes para melhorar a tomada de decisões por meio de decisões guiadas por dados" (BERSIN, 2013).

A contribuição da tecnologia para a melhora da tomada de decisão pode beneficiar toda a companhia, e não somente a área de RH, devido à maior disponibilidade de dados e a maior

capacidade de processá-los. Tradicionalmente, os modelos têm sido criados usando subconjuntos dos dados chamados de amostra. Uma vez que o modelo é criado, um segundo conjunto maior de dados é utilizado para testar ou validar o modelo. Este processo iterativo é frequente. Uma vez que o modelo foi validado, o modelo é utilizado para avaliar ou analisar todos os dados. Esta é uma abordagem estatística bem conhecida que tem sido utilizada durante décadas. Com *Big Data*, cientistas de dados a utilizam uma maior quantidade de dados ou todos os dados para criar o modelo. Ao fazer isto, os cientistas de dados são capazes de introduzir variáveis preditivas no modelo a fim de aumentar a exatidão do modelo (MINELLI, CHAMBERS e DHIRAJ, 2013).

*Big Data* também abre a possibilidade de um modelo de decisão diferente, pois permite às empresas a criação de experimentos controlados para testar hipóteses que guiarão a tomada de decisão em, por exemplo, novos investimentos ou mudanças operacionais. Possibilitando centenas ou milhares de experimentações é possível distinguir entre simples correlação de eventos daqueles que realmente possuem uma ligação de causa e efeito (BROWN, CHUI e MANYIKA, 2011).

Gallant (2011) ressalta a importância da análise de eventos e da capacidade de criar sistemas preditivos. Prever a intenção de um cliente de mudar-se para a concorrência tem muito mais valor do que analisar porque o cliente deixou a empresa. Não que entender os motivos que levaram a esta decisão do cliente não tenha seu valor, mas poder tomar ações de contenção, impedindo a saída do cliente, é mais importante do que agir reativamente, melhorando o que estava errado ou investindo em campanhas caras para tentar trazer o cliente perdido de volta. Esta mesma visão é compartilhada por Hea (2013) que adiciona a importância do uso de dados disponíveis em mídias sociais para a capacidade de realizar estas previsões (NOVO e NEVES, 2013).

Uma terceira maneira pela qual *Big Data* pode contribuir para a melhora de tomada de decisão é evitando o chamado *Big Paradox*. Michael J. Mauboussin apresentou em seu livro "*The Success Equation: Untangling Skill and Luck in Business, Sports, and Investing*", uma "pesquisa de 250 investidores experientes" que ilustra bem como o *Big Data* pode melhorar a tomada de decisão. De acordo com a pesquisa, quase dois terços dos entrevistados concordavam com a afirmação "Conforme uma previsão / recomendação tarefa se torna mais complexa e difícil, eu tendo a confiar mais em julgamento e menos em uma análise formal e quantitativo". Em face da complexidade, os investidores costumam aderir a "um modo mais simples e intuitivo", como escreveu o autor. Ou seja, conforme os problemas se tornam

maiores e mais complexos (juntamente com os dados por trás deles), as pessoas tendem a considerar os dados inúteis e a não querer usar mais usá-los. *Big Data* pode impedir justamente isso por meio da melhora do processo de modelagem dos dados (SIMON, 2013).

### **INOVAÇÃO INCREMENTAL DE PRODUTOS E SERVIÇOS JÁ EXISTENTES**

Um grande número de indústrias - incluindo saúde, o setor público, varejo e indústrias de transformação - pode, obviamente, beneficiar da análise de sua crescente quantidade de dados. A coleta e análise de dados transacionais dá às organizações mais conhecimento sobre as preferências de seus clientes (BUCHOLZ, 2012), de modo que os dados podem então ser utilizados como base para a melhora dos produtos e serviços já ofertados e para a criação de novos produtos e serviços. Isso permite que as organizações possam resolver os problemas emergentes de uma forma oportuna e mais competitiva (OHLHORST, 2013).

Entre as novas fontes de dados, por exemplo, destaca-se a localização de clientes, dados oriundos de *mobile devices* e redes sociais. Dessa maneira, é possível saber não só o que o seu cliente está dizendo sobre você nas mídias sociais, mas exatamente quando ele ou ela entrou em sua loja, permitindo assim que a loja se oriente para melhorar a experiência de seus clientes (DAVENPORT, 2014). De fato, no processo de escolha e compra, os consumidores geram grande volume de dados e a análise técnica desses dados permite a melhora dos serviços prestados e dos produtos ofertados. A análise técnica do processo de compra tornou-se uma característica importante para agências de marketing digital que seguem clientes quando ele ou ela busca informações, compara produtos e, finalmente, toma a decisão de comprar um produto e compra-lo (LEEFLANG, VERHOEF, *et al.*, 2014).

Brown (2011) afirma que as lojas líderes do varejo estão constantemente executando experimentos para identificar fatores que levem a maior participação dos clientes e consequentemente maiores vendas. Embora muitas empresas já façam estes experimentos há algum tempo, o *Big Data* pode levá-los a um próximo nível. A futura geração de lojas de varejo será capaz de identificar comportamentos individualizados de cada cliente, utilizando, por exemplo, o *clickstream*, que é a sequência de cliques de mouse em sua página, determinando o padrão de navegação, se adaptando as preferências do cliente em tempo real (ULARU, PUICAN, *et al.*, 2012). Neste modelo é possível identificar quando um cliente está próximo de efetuar uma compra, permitindo a empresa influenciar na decisão do cliente (NOVO e NEVES, 2013).

Gallant (2011) descreve o caso de uma empresa de telecomunicações da Índia que aproveitou o *Big Data* para melhorar a experiência dos seus usuários. A empresa estava adicionando à sua base uma média de 5 milhões de novos assinantes por mês, entretanto estavam perdendo 1,5 milhões de clientes para a concorrência todos os meses. Após uma análise minuciosa, a empresa identificou que os clientes tinham uma alta tendência de mudar de operadora depois de 6 chamadas interrompidas. Desta forma a empresa, após a quinta chamada interrompida, disparava um torpedo com uma promoção para o cliente, evitando assim a sua fuga para a concorrência. Apenas com um modelo de *Big Data* foi possível a análise de bilhões ou até mesmo trilhões de registros ou eventos, que permitiram a empresa criar um modelo individualizado de relacionamento com seus clientes (GALLANT, 2011).

É interessante ainda chamar a atenção de que a solução da questão de casos como esse só é possível graças à possibilidade, adquirida com *Big Data*, de analisar todo o universo de clientes da companhia. Os modelos tradicionais utilizam amostras randômicas para analisar e obter um quadro geral da situação, o que não é útil na hora da análise de redes específicas de clientes ou de sub-redes de contatos, tendo em vista que considerar estas sub-redes como o universo do qual irá se refletir em uma amostra em amostras demasiadamente pequenas para tirar quaisquer conclusões (BLOEM, VAN DOORN, *et al.*, 2012).

Outro exemplo de melhora da experiência dos usuários é a forma como as estradas de ferro holandesas usam seus próprios dados de viagem e os dados fornecidos por sistemas de navegação para fornecer informações aos clientes sobre seu tempo de viagem previsto pelo trem e pelo carro. Isso ajuda os clientes a tomar decisões mais informadas sobre a sua escolha de viagens (MCAFEE e BRYNJOLFSSON, 2012).

Todos esses exemplos da literatura acadêmica indicam que um dos principais benefícios do *Big Data* é a possibilidade de inovar de maneira incremental produtos e serviços já existentes e criar novas oportunidades de negócios.

### **INOVAÇÃO DISRUPTIVA DE PRODUTOS, SERVIÇOS E MODELOS DE NEGÓCIOS: NOVOS INSIGHTS E DESENVOLVIMENTO DE NOVOS MERCADOS**

Peter Drucker comentou em 1998 que a maioria dos sistemas de informação foram focados em dados contábeis internos e que, como consequência, havia uma tendência degenerativa de gestores, especialmente em grandes corporações, de focar em custos e esforços ao invés de focar em oportunidades externas. Quanto mais interna era a informação obtida pelo *top management*, maior seria a necessidade de balanceá-la com a informação externa – e isso não

existia ainda. Agora, com *Big Data*, o foco externo começa a existir, permitindo a geração de *insights* (DAVENPORT, 2014).

Hoje em dia *Big Data* também é considerado como uma importante fonte de inovações. O desafio aqui é que as empresas têm como objetivo desenvolver novos produtos ou melhoria dos processos de atendimento ao cliente usando várias fontes de dados. Isso faz parte de uma melhoria do processo de inovação, que passa a ser mais *data-driven*. Empresas de cartão de crédito, por exemplo, melhoraram o processo de identificação de fraudes por meio do *Big Data*, por meio da análise do comportamento fraudulento com cartões de crédito roubados. Com base nisso, eles bloqueiam automaticamente cartões de crédito que mostram os padrões de pagamento frequentemente exibidos com cartões de crédito roubados, a fim de reduzir o risco financeiro de tanto o cliente como a companhia de cartão de crédito (ULARU, PUICAN, *et al.*, 2012; MCAFEE e BRYNJOLFSSON, 2012).

O mais ambicioso que uma organização pode fazer com *Big Data* é empregá-lo no desenvolvimento de novos produtos e serviços com base nos dados, isto é, inovar de maneira disruptiva. Algumas empresas têm feito isso como o LinkedIn (desenvolvimento do recurso *People You May Know*, PYMK) e o Google (algoritmo de sugestão de links), entre outras (DAVENPORT, 2014).

Isso tem implicação muito importante, é claro, para o foco da organização de *Big Data* e para os processos e ritmo de desenvolvimento de novos produtos: uma plataforma para a criação de produtos e design tem que ser criação tal qual alguns processos de teste desses novos produtos em uma escala pequena antes que os produtos sejam entregues aos clientes. Obviamente, qualquer um que deseje criar produtos e serviços baseados em dados tem que trabalhar em conjunto com uma equipe de desenvolvimento de produtos, e talvez em conjunto com o marketing. Esses projetos devem provavelmente ser patrocinados pelo líder do negócio em vez de um técnico ou um cientista de dados (DAVENPORT, 2014).

Ter a inovação de produtos e serviços com base em *Big Data* também tem implicações para a avaliação financeira de seus esforços. O desenvolvimento de produtos geralmente é visto como um investimento em vez de oportunidades de economia. Com esse foco, talvez não haja muita economia de tempo ou dinheiro, mas você vai adicionar alguns números grandes ao resultado final da companhia (DAVENPORT, 2014).

Uma empresa global desenvolveu um novo produto, após utilizar o *Big Data* para melhorar o seu processo de manufatura, eles aprenderam tanto sobre os próprios processos que decidiram

prestar consultorias para outras empresas em situações semelhantes. Atualmente tal estabelecimento agrega dados de chão de fábrica e cadeia de suprimentos de diversas empresas e vende ferramentas de software para melhorar a performance destas empresas. Este novo produto/serviço baseado no *Big Data*, disruptivo quando comparado ao modelo de negócios da organização, atualmente representa uma renda maior do que sua manufatura (NOVO e NEVES, 2013).

Por fim, *Big Data* também pode contribuir com a diminuição do tempo necessário para executar uma inovação. Davenport (2014) escreve que esse é o segundo objetivo mais comum para a adoção de ferramentas de *Big Data*. Esse é o objetivo de companhias como a cadeia de lojas de departamentos Macys. Como o CIO descreveu, “a vantagem principal é que o processo de descoberta se torna muito rápido. Um analista pode construir modelos, executá-los, e se ele não gostar de algo, pode modificar o modelo, tudo em apenas um minuto. Esse ciclo geralmente tem oito horas – se o analista puder fazer tudo. A cadeia de eventos é muito mais contínua, o que significa uma maior qualidade de pesquisa” (DAVENPORT, 2014).

### **2.7.3. Oportunidades em Todos os Departamentos de uma Organização**

Por meio do *Big Data*, as organizações podem obter uma visão melhor sobre seus negócios, levando a eficiência em diferentes áreas (DAVENPORT, 2006; GALLANT, 2011; ULARU, PUICAN, *et al.*, 2012; MINELLI, CHAMBERS e DHIRAJ, 2013; SCHMARZO, 2013; DAVENPORT, 2014).

O setor de vendas, por exemplo, pode usar a análise de dados para monitorar o desempenho dos seus funcionários. Seria possível usar smartphones e dispositivos de localização de carro para monitorar como os vendedores realmente gastam seu tempo, gerando uma série de dados interessantes (DAVENPORT, 2014). Outra possível aplicação nesta área é a otimização das atribuições de recursos de vendas, mix de produtos, modelagem de comissões e classificações contábeis, sem falar na possibilidade de otimizar os preços de bens “perecíveis” como mantimentos, lugares de avião e ingressos para shows, melhorando assim a gestão de preços por meio de melhor conhecimento da demanda (SCHMARZO, 2013).

Recursos Humanos tem sido uma das funções menos orientadas a dados no passado, mas os sistemas de informação de RH e análises com base em seus resultados estão começando a mudar isso. Esta função poderia ir a um passo substancial em termos de análise de dados de localização do funcionário e de comunicação (DAVENPORT, 2014). Além disso, o setor pode se beneficiar muito com a identificação das características e comportamentos dos

empregados mais bem-sucedidos e eficazes da organização, para considera-los nos futuros processos de seleção de contratação (SCHMARZO, 2013).

Muitas vezes identificado como responsável por apoiar a tomada de decisões ou por tomar as decisões mais importantes dentro das organizações, o setor de estratégia, tal qual o RH, tem sido historicamente bastante desprovido de dados para a análise dessas decisões. Alguns dados podem ter sido recolhidos, mas raramente de maneira volumosa. Hoje, não usar uma grande quantidade de dados externos na tomada de uma decisão estratégica pode realmente ser uma imperícia (DAVENPORT, 2014).

O setor de TI pode se beneficiar de *Big Data* em duas áreas que poderiam usar mais decisões baseadas em dados: a confiabilidade e a segurança das operações de TI. A primeira para melhorar o desempenho de equipamentos de informática e de obter e manter os clientes. A segunda, para melhorar a segurança em todos os sentidos: em vez de simplesmente responder a violações de segurança, as organizações podem usar *Big Data* para prever onde e como as ameaças de segurança vão surgir, o que é possível através da análise de dados (ULARU, PUICAN, *et al.*, 2012).

A produção industrial e a fabricação também podem se beneficiar do *Big Data*. Não apenas os produtos fabricados, mas também equipamentos de fabricação contêm cada vez mais sensores de produção de dados. Usinagem, solda e dispositivos robóticos podem apresentar um relatório sobre o seu próprio desempenho e necessidade de serviço. Dessa maneira, o monitoramento de máquinas e desvios de processos que podem oferecer indicadores de problemas de qualidade (SCHMARZO, 2013). As operações de fabricação também podem se conectar com grandes cadeias de fornecimento e abastecer-se de dados para garantir uma oferta de insumos contínua, de modo a alimentar fabricação e garantir que apenas a quantidade ideal de bens será fabricada e que o produto final estará sempre à disposição do consumidor (DAVENPORT, 2014). Além disso, por meio de *Big Data*, é possível identificar quais os fornecedores são mais custo-efetivos na entrega de produtos no prazo e sem danos (BLUMENSCHHEIN, 2014). Mesmo no desenvolvimento dos produtos, *Big Data* pode contribuir: as percepções de uso do produto podem ser utilizadas como insumo para acelerar processo de desenvolvimento de novas versões e melhorar a eficácia no lançamento do produto (SCHMARZO, 2013).

O setor de marketing pode se beneficiar por meio da identificação de quais promoções de marketing e campanhas de publicidade são mais eficazes na condução de tráfego de clientes, engajamento e vendas, ou do uso da análise de atribuição para otimizar os objetivos de

marketing, os comportamentos de clientes e dos canais de vendas, e os níveis de estoque e produtos (SCHMARZO, 2013). Os departamentos de marketing podem também desenvolver uma inteligência geoespacial por meio do *Big Data*, o que pode significar anúncios com informações relevantes, contribuindo assim para o aumento de vendas (MINELLI, CHAMBERS e DHIRAJ, 2013). Além disso, o surgimento do conteúdo gerado pelos clientes da Web 2.0 em vários fóruns, grupos de notícia, plataformas de mídia social, e sistemas de *crowdsourcing*, oferece oportunidades para as organizações conhecerem melhor seus clientes (DOAN ET ALII 2011; O'RIELLY 2005).

Mesmo funções como as operações de loja e o atendimento ao cliente podem se beneficiar do *Big Data*: as operações de loja podem otimizar os níveis de inventário dadas as previsões de padrões de compra em conjunto com as informações demográficas locais, o tempo e os dados de eventos, enquanto o atendimento ao cliente pode usar informações de *call-centers*, a fim de obter perfis dos clientes e, assim, aumentar a satisfação por meio da personalização dos serviços (ULARU, PUICAN, *et al.*, 2012; SCHMARZO, 2013).



### **3. FATORES RELEVANTES NA ADOÇÃO E NO USO DE *BIG***

***DATA*** Na maioria dos casos, *Big Data* é um elemento de mudança quando introduzido em uma empresa, e isto inclui questões de escala, armazenamento e design de centros de dados. O rompimento normalmente envolve custos associados a hardware, software, pessoal e apoio, os quais afetam os resultados (OHLHORST, 2013). No entanto, esses não são os únicos fatores relevantes na adoção e uso da tecnologia.

Os fatores de sucesso para adoção de *Big Data* e projetos de análise de dados são similares àqueles de implementações de sistemas de BI (EVERS, 2014). Irei apresentar a seguir alguns dos fatores relevantes na adoção, sem a pretensão, entretanto, de compor uma lista exaustiva tendo em vista que o presente é um estudo exploratório.

#### **3.1. FATORES RELACIONADOS À ESTRATÉGIA**

##### **3.1.1. Alinhamento à Estratégia da Organização**

*Big Data* está gerando novas categorias de empresas que adotam modelos de negócios orientados por dados. Algumas das empresas que integram essas novas categorias já existiam e desempenhavam um papel de intermediação nas cadeias de valor em que se encontravam gerando valiosos dados produzidos pelas transações comerciais (MCGUIRE, MANYIKA e CHUI, 2012). Outras empresas surgiram com a proposta de usufruir e gerar valor a partir da análise de dados, como, por exemplo, o Facebook e o LinkedIn (OHLHORST, 2013).

Apesar dos benefícios, Yan (2013) afirma que a organização precisa de um plano estratégico pois estas implementações podem ser complicadas, idiossincráticas, e disruptivas. Isto ocorre porque a implementação de um sistema de *Big Data* não é um projeto, mas sim um processo. Isto é, pode evoluir dinamicamente e não necessariamente em direções finitas e previsíveis (YEOH e KORONIUS, 2010).

É necessário que a estratégia de implementação e uso do *Big Data* esteja intimamente conectada a estratégia da empresa, e para isso é necessário um extenso planejamento para decidir quais recursos analíticos são necessários e como eles devem ser aplicados para obter o resultado esperado (NOVO e NEVES, 2013). Dessa maneira, o alinhamento com a estratégia da empresa contribui para nortear as ações e sustentar o compromisso de todos, o que é importante para o sucesso na implementação de um sistema de informação (YEOH e KORONIUS, 2010).

As organizações precisam escolher onde vão aplicar *Big Data* e sua análise de dados dentro de seus processos. O escopo deve comunicar de forma clara e específica o que está e o que não está incluído no esforço, além de ser necessário para desenvolver um orçamento acurado e para manter as etapas em andamento. Uma vez que uma organização tenha priorizado uma área, ela deve escolher um objetivo próprio à área por meio de segmentação, criação de melhores alvos, identificação, estabelecimento de recompensas, e assim por diante. Não se pode simplesmente fazer tudo com análise de uma só vez, por isso a escolha de um objetivo é necessária (DAVENPORT, 2014; INSTITUTE FOR ADVANCED ANALYTICS, 2014). Minelli (2013) vai além e afirma que os líderes precisam definir prioridades de negócios e problemas a serem resolvidos, além de definir roteiros mensuráveis, realizáveis e limitados no tempo.

Para Domingues (2013), o maior desafio do *Big Data* é a credibilidade e confiança na tecnologia como solução, bem como no que ela pode fazer para trazer vantagem competitiva para a organização. Com isso em mente, o autor descreve seis passos que uma empresa deve seguir para que a adoção da tecnologia seja um sucesso. Dos seis passos descritos pelo autor, os quatro passos a seguir estão associados ao alinhamento estratégico entre a organização e a adoção do *Big Data*: compreender os negócios e os dados envolvidos, determinar o problema e como os dados podem ajudar, definir expectativas sensatas e manter o objetivo da adoção em mente.

Por fim, outro motivo que torna o alinhamento à estratégia da empresa necessário são as questões futuras. Há desafios desconhecidos que surgirão a cada aumento de escala e desenvolvimento de novas análises. Alguns desses desafios serão intratáveis com as ferramentas e técnicas atuais, e exigirão abordagens criativas e custo-efetivas (KAISLER, ARMOUR, *et al.*, 2013), que se tornam mais factíveis quando a empresa possui os recursos necessários para tais soluções à disposição. A capacidade de reagir rápido às mudanças de condições e de mercado são bem estudadas literatura acadêmica de estratégia, sob a alcunha de *dynamic capability*.

### 3.1.2. Patrocínio da Alta Gestão

Assim como liderança era um fator crítico para o sucesso de programas de análise tradicionais, também o é igualmente relevante para *Big Data* (MCAFEE e BRYNJOLFSSON, 2012; DAVENPORT, 2014). De fato, o papel da alta gestão na difusão de inovação na organização tem sido bem documentado, sobretudo na implementação de sistemas de informação (SOMERS, NELSON e RAGOWSKY, 2001). Sumner (1999), por exemplo, encontrou que o envolvimento do *top management* era fator relevante para a adoção do sistema de informação em seis *business cases* do total de sete estudados.

Com o *Big Data* não se espera que seja diferente, tendo em vista que a decisão e visão do executivo sempre serão essenciais para o negócio. Não só para as decisões gerais que passarão a ser muito mais embasadas em dados (NOVO e NEVES, 2013), como também para endossar o *Big Data* em si, a tomada de decisão baseada em *analytics* e a identificação dos campeões de consumo de análises de dados. O envolvimento da alta gestão é, portanto, importante devido ao apoio e à anuência dos executivos. Sem foco consciente sobre essas atividades, os relatórios visualmente ricos ou modelos estatísticos sofisticados não vão dirigir qualquer valor de negócio real (MINELLI, CHAMBERS e DHIRAJ, 2013).

O envolvimento da alta gestão também é importante porque torna mais fácil o acesso aos recursos operacionais necessários à implementação, tais como financiamento e habilidades profissionais, além de ser importante na superação de barreiras à mudança e na reformulação de paradigmas mentais (YEOH e KORONIUS, 2010). Isso significa não somente remover os obstáculos, mas também criar o ambiente favorável para a análise de dados, tais como suporte organizacional, defesa de causa, estabelecimento de nova estrutura organizacional e definição dos papéis e responsabilidades de cada um dos profissionais envolvidos com o *Big Data*, bem como compreensão da disponibilidade da equipe e de recursos para o projeto (INFOCHIMPS, 2013).

Poon & Wagner (2000) destacam que, para alavancar o patrocínio dos principais executivos, é necessário que haja um patrocinador operacional designado para gerir os detalhes de implementação. Este patrocinador operacional deve ser familiarizado com o estilo de gestão do patrocinador executivo.

Além disto, um dos traços-chave de liderança para *Big Data* parece ser a vontade de patrocinar atividade experimental com dados em grande escala. Os cientistas de dados devem ter um canal aberto e direto de comunicação com a alta direção, tendo em vista que um dos

objetivos da experimentação é eliminar as barreiras à implementação de ideias e ofertas inovadoras e que pode ser necessário um tempo até que o valor da análise de dados seja reconhecido (DAVENPORT, 2014).

## **3.2. FATORES ESTRUTURAIS E CULTURAIS**

### **3.2.1. Design e Processos Organizacionais**

Um dos desafios associados à implementação de uma nova tecnologia em uma organização é sua compatibilidade com os processos de negócios. Isso ocorre devido à diversas questões que podem ser observadas nas organizações, tais como a existência de silos de dados, engessamento dos processos em relação à descoberta e análise de dados, o surgimento de novos papéis na organização e o desenvolvimento de uma nova cultura.

#### **Silos de dados**

De acordo com Yeoh e Koronius (2010), o propósito primário dos sistemas de BI é a integração dos silos de dados por meio do alinhamento entre os processos de negócios com os processos estruturados pelo sistema, de maneira a melhorar o processo de tomada de decisão das empresas.

Um dos desafios enfrentados pela adoção do *Big Data* nas organizações é o compartilhamento de informação entre unidades de negócio, sejam as organizações empresas privadas, governos ou departamentos de um mesmo governo (OHLHORST, 2013; XEXÉO, 2013), sobretudo quando se considera que para os dados serem úteis eles precisam atravessar as fronteiras organizacionais, o que é, geralmente, difícil e causa brigas e disputas internas por poder corporativo (NOVO e NEVES, 2013).

Brown (2011) confirma que os silos de dados são desafios dentro das organizações e apresenta um exemplo de uma empresa europeia de telecomunicações que ao analisar dados interdepartamentais encontrou *insights* diferentes daqueles que eram encontrados por meio de questionários tradicionais, o que desencadeou uma verificação da campanha de marketing para conseguir concluir se os *insights* encontrados eram validados.

A partilha de dados entre os diferentes departamentos e agências do governo é também um desafio. Cada agência ou departamento governamental normalmente tem o seu próprio armazém de informação e reluta muito em compartilhá-lo. Essa falha de comunicação é, por

vezes, o desafio que impede a integração de dados; por exemplo, no Reino Unido, uma coalizão de departamentos de polícia e hospitais destinada a partilhar dados sobre crimes violentos foi relatada como um fracasso devido a uma falta de comunicação entre as organizações participantes (KIM, TRIMI e CHUNG, 2014).

### **Engessamento dos processos em relação à descoberta e análise de dados**

Outro desafio relacionado aos processos organizacionais é a maneira como se lida com a descoberta e a experimentação dos dados. Até o momento, o foco principal das organizações empresariais tem sido a automatização de processos. A análise de dados tem sido usada para compreender e ajustar os processos, mantendo a gestão informada e alertando-os para anomalias. *Big Data* transforma essa abordagem. Empresas de *Big Data* inovam constantemente, explorando e experimentando para saber mais sobre suas operações e os clientes, adotando um forte foco em inovação e exploração, e reagindo rapidamente à constante mudança do universo de negócios em que estão inseridas (DAVENPORT, 2014). Para adotar o *Big Data* com sucesso, as organizações devem, portanto, adotar uma abordagem mais aberta à exploração, experimentação e às tentativas de inovação.

De fato, muitos gestores terão que mudar suas mentalidades e hábitos para que os dados externos e de formato variado sejam incorporados nas tomadas de decisão, nos produtos e nos serviços. Eles vão ter que regularmente examinar fontes externas de dados para ver o que está disponível e como isso pode apoiar a sua organização, entre outras coisas. Um dos impactos mais significativos de *Big Data* é a mudança organizacional ou transformação necessária dos processos para apoiar e explorar a grande oportunidade de dados (DAVENPORT, 2014).

### **Surgimento de novos papéis na organização**

Outro desafio no que tange o alinhamento aos processos organizacionais está relacionado ao surgimento de novos e críticos papéis que precisam ser adicionados à organização na adoção ao *Big Data* (SCHMARZO, 2013).

Entre os novos papéis necessários, Schmarzo (2013) destaca a equipe de experiência do usuário e os novos cargos de gestão sênior. A equipe de experiência do usuário é a responsável por entender as necessidades do usuário, validar os insights oriundos dos dados captados e dar o feedback dos produtos e serviços que surgem a partir da análise de dados.

Os novos cargos de gestão sênior são o *Chief Data Officer* e o *Chief Analytics Officer*. O primeiro é o responsável por adquirir, armazenar, enriquecer e alavancar os ativos de dados da companhia. Isto é, eles são os responsáveis pelas atividades de inventário dos dados,

análise econômica dos dados, monetização dos dados e instrumentação e governança dos dados. O *Chief Analytics Officer*, por outro lado, é o responsável por capturar e rastrear os modelos de análise e os insights resultantes da análise que são desenvolvidos e explorados na organização, o que envolve muitas responsabilidades relacionadas com avaliação e análise de ativos, gestão de propriedade intelectual e proteção, aplicação de patente e monetização da propriedade intelectual (SCHMARZO, 2013).

O surgimento dessas novas funções impacta diretamente o organograma da organização e, conseqüentemente, os processos de negócios. Acomodar essa nova realidade é um desafio para as organizações, sobretudo, devido à cultura já existente na organização. A própria equipe de *Big Data* é um exemplo disso, tendo em vista que ela deve ser multidisciplinar, como será observado na discussão dos próximos candidatos a fatores relevantes na adoção e uso da tecnologia.

### **O desenvolvimento de uma nova cultura**

Embora ainda seja cedo para se referir a uma cultura específica das organizações de *Big Data*, alguns executivos já podem sentir que uma mudança da cultura para facilitar as iniciativas de implementação da tecnologia (DAVENPORT, 2014).

Três mudanças culturais e organizacionais principais são esperadas: maior atratividade pelos dados e os *insights* oriundos de sua exploração, maior cuidado com os modelos de análise – que passam a ser considerados propriedade intelectual –, e o surgimento de um maior conforto na tomada de decisão com base nos dados (SCHMARZO, 2013).

Davenport (2014) vai além e afirma que é esperado que as organizações passem de organizações hierárquicas para organizações meritocráticas, tendo em vista que grandes ideias podem vir de qualquer lugar e de qualquer pessoa na organização. Além disso, de acordo com Davenport (2014), é também esperado o surgimento de uma impaciência como *status quo*, bem como o desenvolvimento de uma cultura de compromisso e arrojada.

Além disso, leva tempo para mudar a cultura de modo que haja maior confiança na análise de dados tradicionais, sobretudo porque o *Big Data* força os funcionários a lidar com novas fontes de dados e novos processos de trabalho, sem falar em quantidades e velocidades de dados são maiores do que nunca. Muitas pessoas têm que sair de suas zonas de conforto. Logo, as organizações devem esperar alguma resistência de seus empregados, especialmente daqueles que acreditam que possuem habilidades e capacidades de julgamentos únicos

(BURTON, MASTRANGELO e SALVADOR, 2014). Com efeito, é necessário que haja uma visão estratégica de longo prazo para *Big Data* (NOVO e NEVES, 2013) para facilitar a percepção do valor agregado pela tecnologia e diminuir a resistência à adoção e uso dela.

### **3.3. FATORES RELACIONADOS AOS RECURSOS HUMANOS**

Um dos elementos mais importantes de um projeto de *Big Data* é quem irá realizar as análises (OHLHORST, 2013). Conforme bem resumiu Minelli et alii (2013), “a tomada de decisão baseada em dados é uma jornada, e sem o talento certo ao longo dos níveis organizacionais, os benefícios da ciência da decisão não podem ser efetivamente realizados”.

#### **3.3.1. Talent Gap**

Um desafio que *Big Data* pode impor às organizações é a falta de profissionais com as habilidades e conhecimentos necessários para lidar com a análise dos dados, o que Leeftang, Verhoeff, Dahlström et alii (2014) definiram como *Talent gap*. Como Davenport (2012) escreveu, “um novo papel tem ganho proeminência nas organizações: o do cientista de dados. Cientista de dados é o profissional que entende como buscar respostas importantes para as questões do setor na enorme quantidade de dados não estruturados que a empresa passa a lidar e armazenar. Conforme as companhias capitalizam o potencial do *Big Data*, a maior é a restrição com que elas se deparam diante da escassez desse talento especial”.

Como ocorreu com a definição de *Big Data*, há diversas definições para o cargo cientista de dados também. Todas as definições indicam que os cientistas de dados precisam adquirir mais competências e realizar mais tarefas do que os estatísticos tradicionais ou analistas quantitativos, porque, são responsáveis por ajudar a desenvolver produtos e serviços voltados para o cliente, a partir de dados desestruturados (YAN, 2013).

Por um lado, os profissionais tem que ir além da análise dos dados tradicionais, pois o conjunto de dados tem muito mais volume e variedade, e deve ser analisado como uma velocidade muito maior – o que tornam necessárias as habilidades substanciais de um funcionário de TI. Por outro, os profissionais também precisam estar perto de produtos e processos internos das organizações, o que significa que eles precisam ser organizados de forma diferente daquela com a qual eram organizados os analistas de dados no passado (DAVENPORT e PATIL, 2012).

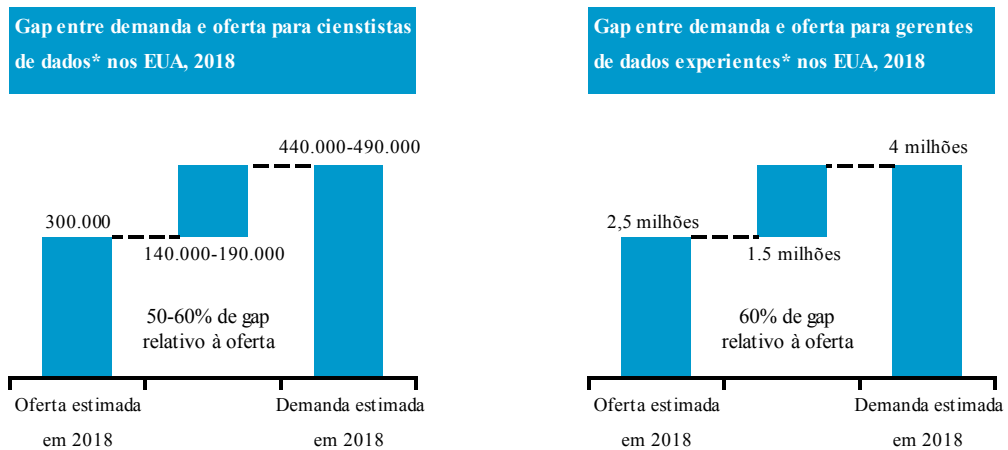
Para Davenport (2014) e Yan (2013), o cientista de dados deve:

- (a) saber codificar e programar além de ter alguma familiaridade com as tecnologias comuns à *Big Data* tal como um hacker;
- (b) ter conhecimento estatístico e de modelagem de dados, pois ele deve extrair a informação dos dados e apresentar a fonte das conclusões sempre que requisitado;
- (c) estar confortável em improvisar, ser capaz de aprender rápido e pelo método do faça-você-mesmo, e ser impaciente e curioso tal como um cientista, pois estes são capazes de formular hipóteses e testá-las com um rigoroso raciocínio científico;
- (d) ter fortes habilidades de relacionamento, visualização de dados e comunicação, tal como os melhores consultores, pois eles devem apresentar seus *insights* de maneira que os executivos entendam;
- (e) ter um conhecimento razoável do mercado e do setor em que se encontra a organização e sobre os processos de negócios usuais da atividade em questão, tal como um especialista em negócios, pois o cientista de dados tem que ser um contador de histórias que pode se comunicar e comunicar suas descobertas de *Big Data* para explicar os padrões e as possíveis relações – o que só é possível se houver um prévio conhecimento da atividade em que o negócio está inserido.

Encontrar o talento certo para analisar dados é um desafio para as organizações, pois as competências exigidas não são simples e nem exclusivamente orientadas para a tecnologia. Procurar e encontrar os cientistas de dados competentes é difícil e caro para a maioria das organizações (KIM, TRIMI e CHUNG, 2014).

Com o interesse crescente em *Big Data*, identificar este indivíduo que tem as habilidades para detectar e extrair valor dos dados tornou-se uma das principais prioridades para muitos executivos. Segundo o McKinsey Global Institute (2011), em 2018, só nos Estados Unidos haverá uma lacuna de 40 a 50 por cento entre a oferta e a demanda necessária de profissionais adequados para trabalhar com a análise de dados, como pode ser visto na figura abaixo (YAN, 2013).

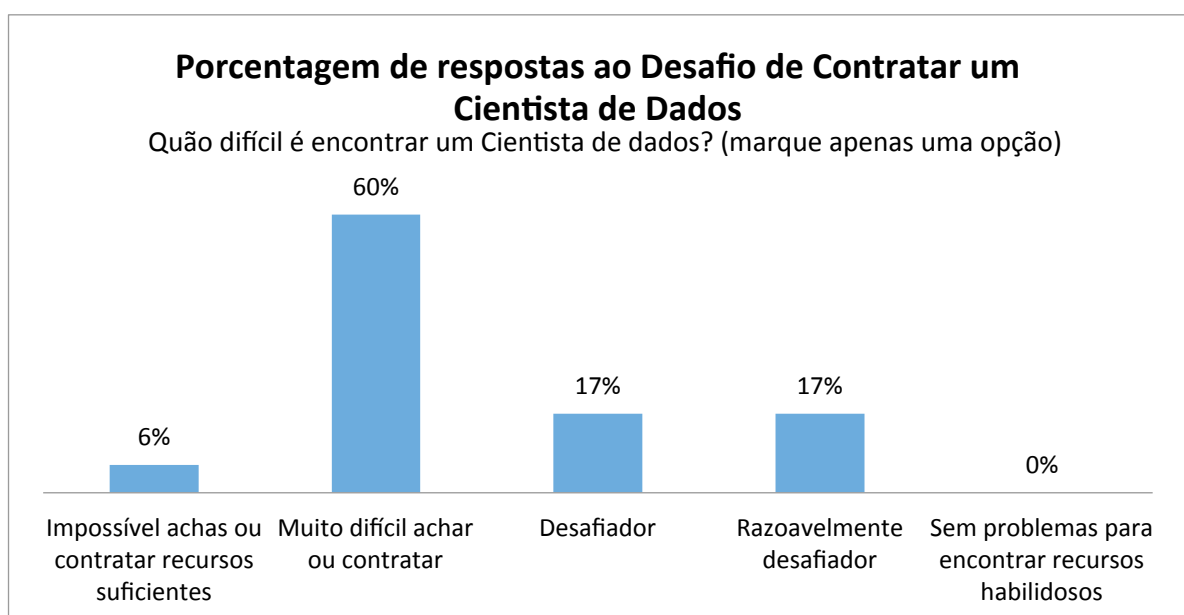




**Figura 8:** Gap entre a Demanda e a Oferta de Cientistas de Dados e Gestores de Dados nos EUA em 2018  
(Fonte: Yan, 2013)

Em outras palavras, isso significa que entre 440 e 490 mil pessoas treinadas para analisar dados serão necessárias nos Estados Unidos em 2018. A oferta provável, no entanto, só será capaz de prover 300 mil desses tipos de funcionários (LEEFLANG, VERHOEF, *et al.*, 2014).

De fato, há uma enorme dificuldade em encontrar alguém com todas as habilidades descritas (DAVENPORT e PATIL, 2012), conforme foi constatado pela *Big Data Executive Survey 2013*: mais de 60 por cento dos entrevistados afirmaram que é muito difícil encontrar ou contratar cientistas de dados para suas organizações (NewVantage Partners, 2013 APUD YAN, 2013), conforme pode ser visto na figura a seguir.



**Figura 9:** Resumo do Relatório *Big Data Executive Survey 2013*

O problema revela-se ainda maior quando se percebe que o talento – o profissional com todas as habilidades necessárias para executar a tarefa – é o recurso mais importante para explorar com sucesso o *Big Data*. Os cientistas de dados extraem os dados a partir de locais diversos, escrevem programas para transformar dados não-estruturados em estruturados, analisam os dados, interpretam os resultados e aconselham executivos sobre o que fazer sobre isso - tudo em curto espaço de tempo e com um senso de urgência. O outro fator complicador para se concentrar nos indivíduos multitalentosos é que os cientistas de dados são apenas um componente de uma abordagem eficaz para projetos de *Big Data* (DAVENPORT, 2014).

### **Soluções propostas para combater o *Talent Gap***

Várias soluções têm sido propostas para resolver a crise de talentos: o investimento na educação formal através de instituições de ensino superior, a formação interior de talentos, a importação de profissionais de países estrangeiros, a terceirização/*outsourcing* e o *crowdsourcing* são algumas das soluções propostas (MILLER, 2013; YAN, 2013).

A educação formal através de instituições de ensino superior como solução para o *Talent gap* consiste na criação de cursos de ensino superior que priorizem habilidades interdisciplinares, entre os cursos de estatística, análise, ciência da computação e matemática (MILLER, 2013). De acordo com os Programas de Pós-Graduação em *Big Data Analytics* e Ciência de Dados nos Estados Unidos, as universidades norte-americanas estão oferecendo ou planejam oferecer cursos de análise avançada de dados, com foco em *Big Data* (GIL PRESS, 2013 APUD YAN, 2013). Algumas universidades como a *Ottawa University*, *North Carolina State University*, e *De Paul University* começaram a oferecer programas formais de graduação em análise de dados (MINELLI, CHAMBERS e DHIRAJ, 2013).

Outra solução é a formação interior de talentos por treinamento. Como o número de diplomados com formação acadêmica formal só é capaz de atender parte da demanda por profissionais, o treinamento interno dos profissionais, naturalmente, torna-se a segunda solução. Uma vantagem desta solução é que os atuais funcionários terão mais conhecimento do domínio da empresa do que um recém-chegado. Essa solução já foi amplamente utilizada com sucesso para atender desafios semelhantes no processo de adoção e implementação de outros sistemas de informação, que tiveram amplo interesse do mercado (SOMERS, NELSON e RAGOWSKY, 2001; UMBLE e UMBLE, 2002).

A opção pelo desenvolvimento interno tem sido bem adotada pelas empresas que se destacam no mercado de *Big Data*. A EMC<sup>2</sup>, por exemplo, ao concluir que a disponibilidade de

cientistas de dados seria um fator decisivo importante nos seus próprios projetos e nos projetos de seus clientes, adquiriu a Greenplum – empresa de destaque em análise de dados desestruturados e em grande volume – em 2010 para expandir suas capacidades em ciência de dados e em seguida criou um programa de treinamento para seus funcionários e clientes. Outras empresas, como, por exemplo, a Deloitte, têm trabalhado com escolas de negócio, como a *Kelley School of Business*, da Universidade de Indiana, para desenvolver habilidades de análise de *Big Data* em seus funcionários (DAVENPORT, 2014).

Há ainda empresas como a Mu Sigma, que investiu em uma alternativa mista entre buscar profissionais formados formalmente e o treinamento interno ao criar uma universidade para suprir a necessidade de talento em seu negócio global (MINELLI, CHAMBERS e DHIRAJ, 2013).

Outra solução possível para suprir a companhia com profissionais dotados dos talentos necessários é a busca por talentos externos, o *outsourcing* e a terceirização. Nesse sentido, a importação de mão-de-obra de países estrangeiros e terceirização são duas outras soluções possíveis para adicionar abastecimento do mercado (MANYIKA, CHUI, *et al.*, 2011). A tarefa de análise de grandes quantidades de dados é desafiadora o suficiente mesmo para os líderes de análise de dados que têm buscado talentos externos para melhorar suas capacitações e competências. O Netflix, serviço de *streaming* de séries e filmes, criou um prêmio para aqueles que conseguissem melhorar os algoritmos de recomendação até então utilizado pela organização, com o intuito de obter os melhores analistas de dados trabalhando na melhora de seu algoritmo sem ter que contratar novos funcionários (SHMUELI, 2012 APUD DHOLAKIA e DHOLAKIA, 2013).

A busca por soluções externas, no entanto, não pode ser considerada uma solução definitiva. Conforme as capacitações analíticas se tornam importantes estrategicamente para as companhias, o *outsourcing* completo das habilidades de análise de dados pode se tornar uma estratégia perigosa. É recomendável para a maioria das companhias manter certo nível de investimento em suas próprias capacitações, contratando empregados (DAVENPORT e PATIL, 2012). Quase duas décadas antes, Slaughter & Aang (1996) afirmaram o mesmo para sistemas de informação anteriores ao *Big Data*, acrescentando ainda a necessidade por flexibilidade e o equilíbrio entre a oferta e a demanda da habilidade técnica como outras dimensões importantes na hora de decidir entre desenvolver internamente as habilidades necessárias por meio do treinamento dos empregados ou buscar soluções de *outsourcing*.

Um outro meio de terceirização dos profissionais de talentos é pela contratação de cientistas de dados como consultores. Nos últimos meses, tem ocorrido um aumento na demanda por cientistas de dados de organizações de consultoria (DAVENPORT, 2014). Enquanto no passado capacitações sofisticadas eram oferecidas por agências de pesquisa de mercado e provedores de dados, como AC Nielsen, GfK e Experian, hoje é possível observar as principais companhias de consultoria oferecendo soluções analíticas. Além disso, agências menores também oferecem essa solução. Um exemplo particular é a DunnHumby que se tornou famosa por seu trabalho junto a gigante Tesco do Reino Unido e agora atende a grandes clientes globais das indústrias do varejo e de alimentação (HUMBY ET AL., 2008 APUD LEEFLANG, VERHOEF, *et al.*, 2014).

Ainda no sentido de buscar uma solução externa, outra solução inovadora é *crowdsourcing*, onde a necessidade de serviços, ideias ou de conteúdo são obtidas através de solicitação de contribuições de um grande grupo de especialistas, especialmente a partir de uma comunidade online, ao invés de empregados ou fornecedores tradicionais (YAN, 2013).

### **3.3.2. Formação de Equipe**

Uma outra alternativa para solucionar a necessidade de talentos é a montagem de uma equipe cujos funcionários tenham as habilidades necessárias em conjunto, sem que cada um dos funcionários tenha necessariamente que ter todas as habilidades (DEVLIN, ROGERS e MYERS, 2012; DHOLAKIA e DHOLAKIA, 2013).

Quando a solução utilizada é a formação de uma equipe, cientistas de dados não são a única preocupação na formação de uma equipe de análise de dados; há também um impacto importante na tomada de decisão por parte dos gestores e executivos. Tão importante quanto contratar os integrantes da equipe é a necessidade de identificar um gerente de projeto e especificar papéis e responsabilidades de cada membro da equipe. Os gerentes da equipe de análise de dados também desempenham um papel de liderança e de construção de cultura no que diz respeito à *Big Data* (DAVENPORT, 2014).

Dada a diversidade de habilidades necessárias, os gerentes de cientistas de dados também terão de ser bons em montar equipes com as habilidades necessárias para cada projeto. Alguns projetos exigirão mais recursos de dados, outros, mais habilidades analíticas e assim por diante. O gerente de ciência de dados terá de determinar as necessidades específicas do projeto antes de montar uma equipe (DAVENPORT, 2014).

Os requisitos para pessoal de apoio da equipe de *Big Data*, por exemplo, são diferentes. Enquanto alguns podem ter habilidades específicas que contribuam com o aumento de produtividade – como, por exemplo, analistas de bancos de dados – outros precisam ir além das habilidades substanciais de TI e possuírem mais visão de negócios e habilidades criativas. Além disso, a equipe de *Big Data* precisa estar perto de produtos e processos dentro das organizações (DAVENPORT, 2014).

A abordagem por equipe foi amplamente utilizada em projetos de implementação de sistemas de TI no passado, como sistemas de BI, ERP, *Knowledge Management* (KM), entre outros. Dessas experiências passadas, fica o aprendizado de que a qualidade da equipe que dá suporte a um sistema de informação é importante. O gerente de projeto deve ter técnica e a capacidade de se comunicar com a gerência sênior. O pessoal de apoio deve ter habilidade para interagir com a gestão de topo da organização e ser capaz de dominar as tecnologias necessárias (POON e WAGNER, 2001).

Yeoh & Koronius (2010) concordam que a composição e as competências de uma equipe têm uma grande influência sobre o sucesso da implementação de sistemas. Segundo os autores, a equipe deve ser multifuncional e composta por ambos os técnicos e de negócios, tendo em vista que a equipe tem que lidar com diversas plataformas e interfaces, conexões com sistemas legados, diversas ferramentas, em suma atividades que exigem diferentes habilidades e competências. Um mix adequado de conhecimentos técnicos e de negócios é a chave para o sucesso (YEOH e KORONIUS, 2010).

Além disto, vale lembrar que um projeto de *Big Data* pode ter mais de uma estrutura dentro da organização. E isso impacta diretamente a maneira como a equipe será formada. De acordo com Minelli, Chambers e Dhiraj (2013), as seguintes estruturas podem ser observadas em muitas organizações: estrutura de análise centralizada, estrutura de análise descentralizada e modelo de federação.

A estrutura de análise centralizada é composta por uma única equipe que oferta os serviços de análise de dados. Enquanto isso carrega a promessa de uma infraestrutura de dados integrada e economias de escala, um modelo centralizado pode não ser capaz de fornecer a agilidade e flexibilidade que é necessário para manter a análise de dados relevante dentro da empresa (MINELLI, CHAMBERS e DHIRAJ, 2013).

A estrutura de análise descentralizada é aquela em que cada função de negócio da organização possui sua própria equipe de infraestrutura análise de dados. Enquanto isso garante a agilidade e flexibilidade, este modelo corre o risco de criar diferentes funções dentro de uma organização que optam por adotar seus respectivos conjuntos de ferramentas e metodologias. Embora esta abordagem alavanque rapidamente cada função, corre-se o risco de criar redundâncias ou pior, abordagens conflitantes, resultando em silos (MINELLI, CHAMBERS e DHIRAJ, 2013).

O modelo de federação procura conciliar as vantagens de ambos os modelos, centralizado e descentralizado. Enquanto cada função permite a flexibilidade de implementação de análise de dados, um conselho de administração assegura que existe um amplo alinhamento em políticas de dados e infraestrutura. Não é necessário dizer que este modelo é extremamente difícil de executar e requer um compromisso dos líderes (MINELLI, CHAMBERS e DHIRAJ, 2013).

	<b>Centralizado</b>	<b>Decentralizado</b>	<b>Federação</b>
Descrição	Organização central de serviços compartilhados que serve múltiplos departamentos	Cada departamento com sua própria unidade de análise de dados	Coordenação central com execução local
Líder	Principais executivos	Líderes das unidades de negócio	Principais executivos como equipes de tenentes incorporados nas unidades de negócios
Prós	Economias de escala em infraestrutura e processos Facilidade de promoção da visão corporativa sobre análise em serviços de capacitações estratégicas. Colaboração entre funções	Falta de alinhamento entre as unidades de negócios Risco de pouca propagação em áreas que precisam de foco profundidade	Mantém a agilidade Acesso direto e profundo às habilidades necessárias Autonomia das unidades funcionais mantidas
Contras	Visualização como centro de custo Falta de alinhamento das unidades de negócios Falta de velocidade e flexibilidade	Abordagem local e de silo Curto prazo e foco Falta de controle e alinhamento com a visão corporativa	Coordenação e planejamento são mais fáceis ditos do que feitos

**Tabela 4:** Pontos positivos e negativos dos diferentes tipos de estrutura (Fonte: Minelli, Chambers e Dhiraj, 2013)

Independente da estrutura adotada, a equipe de análise de dados, bem como os próprios dados por ela acessados, não deve ser de natureza estática e deve ser capaz de evoluir e adaptar-se às necessidades do negócio. Além disso, na formação da equipe em si, as organizações não deveriam contratar apenas qualquer um que mostre uma centelha de interesse ou uma compreensão básica de análise de dados. É importante desenvolver um teste decisivo para determinar se um indivíduo tem as competências necessárias para ter sucesso no que pode ser uma nova carreira. Os candidatos devem possuir uma base de cinco habilidades críticas para trazer imediatamente valor a uma equipe de *Big Data*: (1) a mineração de dados, (2) visualização de dados, (3) a análise dos dados, (4) a manipulação de dados, e (5) a descoberta de dados (OHLHORST, 2013).

Por fim, vale destacar que tão desafiador como é manter uma equipe, reter os bons profissionais também o é. Dada a forte necessidade de estímulo intelectual e crescimento, no entanto, a forma mais importante de manter um cientista de dados é fornecer-lhe com bons dados e problemas interessantes para resolver (DAVENPORT, 2014). Neste contexto, surgem as seguintes proposições:

### **3.3.3. Gestão de Expectativas**

Outro fator relacionado aos fatores humanos é a gestão de expectativas, afinal, as expectativas dos gestores podem exceder as capacitações do sistema (SOMERS, NELSON e RAGOWSKY, 2001). Expectativa é o conjunto de atributos ou características dos serviços ou produtos de informação considerados razoáveis por um usuário; este conjunto é oriundo das funções prestadas pelos sistemas de informação baseados dentro de sua organização (LI, 1997).

“A noção do fracasso das expectativas denota o fracasso do sistema de informação com a falta de habilidade do sistema em atender os requerimentos, expectativas ou valores dos *stakeholders*. Fracasso, portanto, não significa apenas a incapacidade do sistema em atender às especificações técnicas desenhadas” (LYYTIEN e HIRSCHHEIM, 1987). O fracasso das expectativas é percebido como a diferença entre o que de fato é e a situação que era desejada por membros de um grupo de *stakeholders* em particular (YEO, 2002).

O gerenciamento com sucesso das expectativas do usuário tem sido relacionado à implementação com sucesso de sistemas de informação anteriores ao *Big Data* (SOMERS, NELSON e RAGOWSKY, 2001). Para Lim *et. al.* (2005), uma melhor



participação do usuário no processo de mudança pode levar a uma melhor comunicação de suas necessidades e das capacidades do sistema, o que pode ajudar a garantir o sucesso na introdução de um sistema, tendo em vista que a participação formal do usuário pode ajudar a atender as demandas e gerir as expectativas. É importante destacar que a importância do gerenciamento de expectativas é verificada em todas as fases do processo de implementação de sistemas anteriores (SOMERS & NELSON, 2011).

Em relação ao *Big Data*, há sobre este, expectativas muito elevadas, o que favorece ao desapontamento. Desta maneira, é importante considerar que existirão dificuldades de performance na utilização das novas tecnologias. Por causa disso, recomenda-se um longo período de experimentação, que inclua também testes independentes daqueles realizados por indivíduos que vão operar o sistema, bem como envolvimento dos responsáveis pela implementação nas discussões dentro da comunidade de *Big Data* (FREEMAN, 2013).

No momento em que uma organização alcançar um elevado nível de certeza sobre os insights e as implicações, muito mais dados se tornarão disponíveis. Por isso, muitas organizações devem adotar uma abordagem para análise de dados e tomada de decisões mais indicativas e contínuas (DAVENPORT, 2014).

### **3.4. FATORES RELEVANTES RELACIONADOS À GESTÃO DO PROGRAMA DE IMPLEMENTAÇÃO**

#### **3.4.1. Comitê Gestor, Grupo de Superusuários e Campeão de Projeto**

Observado a implementação de sistemas de informação anteriores ao *Big Data*, tais como sistemas de BI, uma característica que se destaca é a criação de um comitê gestor ou um grupo de “superusuários” (SUMNER, 1999), prática que foi reconhecida como importante em todas as seis fases de implementação de um sistema de informação – iniciação, adoção, adaptação, aceitação, rotina e infusão (SOMERS, NELSON e RAGOWSKY, 2001).

Um comitê gestor garante que existam controles adequados sobre os processos de tomada de decisão da equipe e é um meio eficaz de assegurar a participação adequada

dos principais agentes relacionados ao sucesso do sistema, bem como permite que a alta administração monitore o progresso do projeto (SOMERS, NELSON e RAGOWSKY, 2001). No caso dos grupos de superusuários, os benefícios são semelhantes, embora o principal benefício seja o retorno às funções normais ao final da implementação, trabalhando ao lado dos usuários finais e garantindo assim uma maior eficácia do treinamento (SKOK e LEGGE, 2001).

De fato, existem diversos modelos apropriados de gestão e liderança do processo de implementação do *Big Data*. Pode ser estabelecido, por exemplo, um comitê gestor para “facilitar a colaboração dos dados e das capacitações analíticas entre as unidades de negócios” (GROVES, KAYYALI, *et al.*, 2013).

O Comitê Gestor de projetos deve ser interorganizacional, liderado por um executivo sênior e envolvendo funcionários e líderes que não foram diretamente afetadas pela reorganização, de modo a promover dados como um ativo estratégico que pode ajudar a crescer a empresa e torná-lo mais bem sucedido (BURTON, MASTRANGELO e SALVADOR, 2014).

Outra característica que se destaca é o estabelecimento de um campeão de projeto, embora em diferente grau de destaque e importância (SOMERS, NELSON e RAGOWSKY, 2001). É necessário ressaltar que um campeão de projeto não é o mesmo que um grupo de usuários ou um comitê gestor. O campeão deve agir como um defensor do sistema e de seus benefícios (SHANKS, PARR, *et al.*, 2000). As habilidades de liderança transformacional do campeão de projeto desempenham um papel fundamental no sucesso da implementação, ao resolver conflitos e gerenciar resistência (STEFANOU, 1999) e gerir a mudança (MURRAY e COFFIN, 2001). Campeões de projeto desempenham um papel crítico na aceitação da tecnologia e, em menor grau, durante a sua utilização e incorporação na organização (SOMERS, NELSON e RAGOWSKY, 2001), pois tem excelente visão de negócios, e são capazes de prever os desafios organizacionais e mudança de curso em conformidade (YEOH e KORONIUS, 2010).

A execução de um projeto de análise de dados não é diferente. Antes de executar um projeto como esse, a organização deve avaliar o valor total de negócio, considerando não só o aspecto financeiro, mas também verificando a facilidade de uso da solução criada, a escalabilidade do projeto, a criação de padrões e a prontidão da empresa para aportar o projeto. E para tal, a alta administração deve conhecer o interesse dos diversos

níveis organizacionais – inclusive o das várias divisões de TI – no projeto. No mínimo, deve ser estabelecido um patrocínio executivo, um comitê gestor ou um campeão do projeto, responsável pelas decisões técnicas, e um tomador de decisão econômica (INFOCHIMPS, 2013).

Isto é, como todos os projetos de implementação de sucesso, também em *Big Data* é importante identificar o responsável por remover os obstáculos, gerenciar o orçamento, fornecer suporte organizacional, identificar as áreas que vão render recompensas (de eficiência, de qualidade ou financeiras), identificar como incentivar cada um dentro da organização e, por fim, transformar a visão do uso da análise de dados em um esforço da comunidade. Isto é, servir como um campeão de projeto (INSTITUTE FOR ADVANCED ANALYTICS, 2014; PATTERSON, 2014; LEININGER, 2015)

### **3.4.2. Comunicação**

Outro aspecto que se destacou como fator relevante para a adoção e implementação de sistemas de análise de dados anteriores ao *Big Data* foi a comunicação, que inclui desde o ato de promover formalmente as equipes de projeto até a comunicação do andamento do projeto para o resto da organização e gestão de expectativas (HOLLAND, LIGHT e GIBSON, 1999). A comunicação revelou-se importante para a melhor compreensão e entendimento do escopo do projeto (SUMNER, 1999; NAH, LAU e KUANG, 2001), uma melhor educação sobre os novos processos de negócios, cooperação entre os diversos departamentos da empresa (SOMERS, NELSON e RAGOWSKY, 2001; ROBINSON e DILTS, 1999), e a resolução de conflitos (LI, 1997; HOLLAND, LIGHT e GIBSON, 1999).

De fato, um dos primeiros passos para determinar se um projeto terá sucesso é definir objetivos de negócio claros e critério de sucesso específicos, que sejam compartilhados e aprovados pelas principais partes interessadas do projeto, o que demanda uma comunicação eficiente com todos os envolvidos no projeto e todos aqueles que serão impactados pelo projeto (INSTITUTE FOR ADVANCED ANALYTICS, 2014). Somente dessa maneira todas as partes podem desempenhar seu papel tão bem quanto necessário para o sucesso do projeto.

A contínua cooperação entre os departamentos e a administração de problemas são essenciais para o sucesso dos projetos de *Big Data*. Os conhecimentos adquiridos a partir de um projeto de análise de dados são amplos e podem contribuir para a melhora

do desempenho de diferentes partes, incluindo as equipes de TI, cientistas de dados e analistas, e as pessoas em toda a organização a partir da linha de gerentes de negócios. Linhas de comunicação abertas são necessários para manter os projetos no caminho certo e para entregar valor de negócio (INSTITUTE FOR ADVANCED ANALYTICS, 2014).

Além disso, a comunicação contínua pode melhorar o desempenho ao longo dos diversos projetos de *Big Data* da organização. Projetos de análise de dados nesse porte são mais como projetos de P&D do que como aplicações de produção e, portanto, têm um risco maior de não alinhamento com as expectativas originais ou não. A coleta e comunicação das lições aprendidas a partir de cada projeto ou iteração aumentará a mudança para o sucesso em projetos futuros (INSTITUTE FOR ADVANCED ANALYTICS, 2014).

Um exemplo de como a comunicação contribuiu para o sucesso do projeto de *Big Data* foi a reorganização da RTI International, empresa de pesquisa sem fins lucrativos, que incluiu uma comunicação robusta e um plano de gestão da mudança. Todas as partes interessadas, tais como a câmara municipal da cidade onde o projeto era executado, pequenos grupos relacionados com o problema e membros da equipe ajudaram a desenvolver e revisar as respostas às perguntas mais frequentes. O objetivo na adoção da comunicação massiva foi duplo: alinhar mais de perto a questão a ser resolvida e a expertise em ciência de dados, e sustentar às colaborações dentro das áreas específicas da organização (BURTON, MASTRANGELO e SALVADOR, 2014).

### **3.4.3. Programa de Implementação**

Relata-se que os projetos de *Big Data* falham 30% mais frequentemente do que outros projetos de TI, e as principais razões relatadas para tal são definição imprecisa de escopo, falta de um conjunto de objetivos e expansão irrestrita (INSTITUTE FOR ADVANCED ANALYTICS, 2014). De uma maneira ou de outra, essas razões estão intimamente relacionadas à gestão do Programa de Implementação do *Big Data*.

Isso não é uma surpresa quando se considera que a gestão do Programa de Implementação é um dos fatores relevantes na implementação e adoção dos sistemas de informação anteriores ao *Big Data*, tais como sistemas de BI e ERP (SOMERS, NELSON e RAGOWSKY, 2001). O programa de implementação em si tem diversas dimensões. Vou abordar aqui os dois aspectos mais relevantes a serem considerados

quando da implementação: o método de gestão de projeto e o uso de abordagem faseada e evolucionária.

O primeiro aspecto está relacionado à própria gestão de sistemas e tecnologias da informação dentro das empresas e envolve sobretudo o método de implementação do projeto.

De maneira geral, um projeto de tecnologia da informação pode ser desenvolvido e implementado por diversos métodos, dentre os quais se destacam o método ágil e o método tradicional. Embora sua classificação como relevante para a adoção de tecnologias da informação anteriores ao *Big Data* tenha sido questionada (ANSARINEJAD, AMALNICK, *et al.*, 2011), este não parece ser o caso quando o assunto é *Big Data*. Como o treinamento e educação substancial é susceptível de ser exigido da parte de todos, e requisitos de negócios detalhados podem mudar à medida que há avanços, metodologias de desenvolvimento ágeis se ajustam melhor às grandes aplicações analíticas de dados quando comparadas às abordagens convencionais (SHERMAN, 2005 APUD BORKOVICH e MORRIS, 2011).

A incorporação da impaciência ao *status quo* da organização, referida anteriormente, está também relacionada à adoção de novas metodologias para a visão e desenvolvimento de produtos à base de dados. Métodos tradicionais de Cascata – abordagens altamente estruturadas que produzem um resultado apenas ao fim de um longo processo – têm sido cada vez mais forçadas a sair do sistema de processos de desenvolvimento em favor dos processos mais rápidos e mais flexíveis. Abordagens ágeis, em que relativamente pouco tempo é gasto na especificação prévia de um sistema e mais ênfase é colocada sobre a rápida entrega de pequenos resultados iterativos e que acomodam mudanças rápidas nos planos de desenvolvimento, funcionando melhor em meio a toda incerteza gerada pela nova tecnologia (DAVENPORT, 2014).

Em outras palavras, projetos de *Big Data* são mais adequados para um processo de desenvolvimento ágil e interativo, pois os ciclos de tempo curtos, com resultados rápidos e envolvimento constante do usuário, entregando gradativamente uma solução de negócios. Um método de gestão da implementação tradicional não permitiria um número suficiente de iterações (FAYET, 2015).

O uso de abordagem faseada e evolucionário, é constantemente destacada como um fator crítico de sucesso para a adoção de sistemas de informação (CLIFFE, 1999;

POON e WAGNER, 2001; GUPTA, GUPTA e SINGHAL, 2014). Mandal & Gunasekaran (2003), por exemplo, defendem a divisão do projeto em fases naturais ou subsistemas de planejamento modular para o desenvolvimento das comunicações interfuncionais e que as empresas devem considerar uma implementação gradual em vez de uma abordagem radical (SIRIGINIDI, 2000 APUD FINNEY e CORBETT, 2007).

No caso específico do *Big Data*, a abordagem faseada revela-se importante por causa da necessidade de adaptação à mudanças, que podem ser decorrentes tanto de forças externas como de forças internas e, em geral, estão associadas a diferentes aspectos de esforço necessários e tardiamente descobertos ou tarefas essenciais despercebidas inicialmente, ambos precisando ser adicionados para que o projeto seja bem sucedido (INSTITUTE FOR ADVANCED ANALYTICS, 2014).

Intimamente relacionada com o Programa de Gestão da Implementação e com a abordagem faseada de implementação de tecnologias, a possibilidade de correção de erros das versões iniciais e de resolução de conflitos também é importante quando o assunto é a implementação de um sistema de informação. Embora não tenha tido a mesma importância quando avaliada pelos mais diversos *stakeholders* envolvidos com a implementação de sistemas de informação anteriores ao *Big Data* (FANG e PATRECIA, 2005), a possibilidade de correção de erros de versões iniciais e a resolução de conflitos foram aspectos considerados críticos na implementação de diversos sistemas de informação tais como ERP, CRM, KM e BI (BINGI, SHARMA e GODLA, 1999; HOLLAND, LIGHT e GIBSON, 1999; FANG e PATRECIA, 2005).

Também no caso do *Big Data*, esse aspecto é relevante para a adoção. A prova de conceito é desenvolvida para ajudar os colaboradores e os usuários a terem noção do resultado que está sendo desenvolvido com a implementação da nova tecnologia (PATTERSON, 2014). Patterson (2014) e Leininger (2015), aliás, descrevem passo-a-passo de como obter valor de um projeto de análise de dados e desenvolver uma prova de conceito, de modo a maximizar a possibilidade de sucesso do projeto.

#### **3.4.4. Monitoramento e Avaliação de Performance**

A Gestão do Programa de Implementação de um sistema de informação também deve estar atenta à performance para que implementação seja um sucesso, tendo em vista que

esse foi um aspecto que se destacou como relevante no processo de adoção de sistemas de informação anteriores ao *Big Data* (FANG e PATRECIA, 2005).

O que pode ser apurado nessa literatura é que o acompanhamento do progresso da implementação só é possível se o monitoramento do desempenho avaliar objetivos e metas que sejam mensuráveis e razoáveis (HOLLAND, LIGHT e GIBSON, 1999) e estejam atrelados aos objetivos do projeto. As medidas devem também ser cuidadosamente construídas porque além de indicar como o sistema está sendo executado, elas devem incentivar os comportamentos desejados por todas as funções e indivíduos envolvidos com a implementação (UMBLE e UMBLE, 2002).

Nesse contexto, no intuito de acompanhar o progresso da implementação de sistemas de informação, Barrar & Roberts (1992) desenvolveram dois tipos de critérios para o acompanhamento do progresso: os critérios de gerenciamento de projetos e os critérios operacionais. Os critérios de gerenciamento de projetos são usados para medir os custos de realização e a qualidade obtida levando em comparação as datas de conclusão de cada estágio. Os critérios operacionais são usados para medir as mesmas características, mas levando em comparação o sistema de produção.

*Big Data* não foge à norma dos demais sistemas de informação. O sucesso tem de ser medido, e medir a contribuição de uma equipe para os resultados pode ser um processo difícil. É por isso que é importante para construir objetivos, medidas e metas que demonstrem os benefícios de uma equipe focada em *Big Data Analytics* (OHLHORST, 2013).

Um dos primeiros passos para determinar se um projeto está no caminho do sucesso é definir critérios de sucesso específicos. Estes critérios devem ser mensuráveis com métricas-chave de desempenho, associadas a objetivos claros e documentados de forma tão precisa quanto possível para assegurar sua comunicação correta. Cada objetivo empresa precisa ter critérios mensuráveis que irão permitir determinar a posteriori se o objetivo foi cumprido com êxito (INSTITUTE FOR ADVANCED ANALYTICS, 2014).

O monitoramento da performance é algo tão importante que não envolve apenas um fator crítico de sucesso e sim a seis distintos fatores que estão associados a duas categorias: a performance tecnológica e a performance dos processos e da organização. Na categoria de performance da tecnologia, os fatores críticos são a qualidade do

sistema, a qualidade da informação, a velocidade certa de acesso à informação; na categoria de performance dos processos e da organização, os fatores críticos são o uso da tecnologia (aceitação pelo usuário), custo e tempo na categoria de performance da organização e dos processos (EVERS, 2014). Diversos autores escreveram sobre esses fatores, tais como Yeoh & Koronius (2010), entre outros.

É importante ressaltar que todos esses aspectos, que aqui se resumem à performance e monitoramento, estão associados à definição de *Big Data* adotada pela Gartner Inc. Isto é, estão relacionados à custo-efetividade dos sistemas capazes de lidar com conjuntos de dados de grande volume, velocidade e variedade (CEARLEY e CLAUNCH, 2012).

### **3.5. FATORES RELACIONADOS À GOVERNANÇA DA INFORMAÇÃO**

*“A tecnologia não é boa nem é má, nem é neutra”. Melvin Kranzberg*

#### **3.5.1. Ética**

*Big Data* levanta simultaneamente questões preocupantes para a epistemologia e para a ética, como, por exemplo, quais dados podem ser utilizados em uma análise. Os formuladores de políticas têm lutado muito para definir o que é ético no uso de dados. A discussão historicamente girava em torno da definição de "dados sensíveis". Ainda assim, qualquer tentativa de definir exaustivamente categorias de sensibilidade geralmente falhava, dada a natureza altamente contextual de informações pessoais (TENE e POLONETSKY, 2013).

As questões epistemológicas e éticas, no entanto, não são as únicas a serem discutidas quando o assunto é *Big Data*. Diversas outras questões também são importantes: sob que condições alguém pode/deve ser considerado como parte de um grande conjunto de dados? E se alguma postagem no domínio 'público' é tomada fora do contexto e analisada de uma forma que o autor nunca imaginou? O que significa para alguém ser identificado, ou ser analisado sem o saber? (BOYD e CRAWFORD, 2012).

Só porque o conteúdo é acessível ao público, não significa que ele foi concebido para ser utilizado por qualquer pessoa. Há questões sérias envolvidas na ética da coleta e



análise de dados on-line (ESS, 2002). O processo de avaliação da ética em uma pesquisa não pode ser ignorado simplesmente porque os dados são aparentemente públicos.

Isto ganha importância se considerarmos que os usuários não são necessariamente conscientes de todos os usos múltiplos e outros ganhos que vêm de informações que eles postaram. Os dados podem ser públicos (ou semi-públicos), mas isso não equivale, de forma simplista, a uma permissão total para serem utilizados para todos os fins. E isso se torna importante à medida que analistas de *Big Data* raramente reconhecem que há uma diferença considerável entre estar público e ser público (BOYD e MARWICK, 2011).

Nesse sentido, como a natureza efêmera dos dados pessoais torna difícil a recaptura depois que eles são expostos em uma esfera pública ou semi-pública, faz-se necessário discutir quando e quais dados podem ser considerados como parte da estratégia de *Big Data*, tendo em vista que a dificuldade de garantir a segurança e privacidade de dados chegam a inviabilizar projetos, como uma base central de prontuários médicos, devido ao risco de essa informação ser utilizada de forma indevida. Mesmo uma aplicação altamente benéfica, como o prontuário médico integrado, nas mãos erradas, pode resultar em políticas discriminatórias. É essencial um questionamento ético constante não só sobre o uso, mas também sobre a coleta, o armazenamento e controle de acesso a esses dados (SIMON, 2013).

A fim de delimitar a zona de análise ética de dados, seria importante que as organizações revelassem não só a existência de suas bases de dados, mas também os critérios utilizados em seus processos de tomada de decisão, sujeitos a proteção dos segredos comerciais e outras leis de propriedade intelectual (TENE e POLONETSKY, 2013). Boyd & Crawford (2012) apelam por uma maior responsabilidade por parte dos pesquisadores e sugerem uma reflexão sobre a importância da prestação de contas, para garantir a ética desde a obtenção dos dados até sua análise e assegurar a proteção das pessoas.

Outro ponto importante em relação ao *Big Data* concerne à análise preditiva. A análise preditiva é particularmente problemática quando baseada em dados categorizados como sensíveis, tais quais saúde, raça ou sexualidade. Uma coisa é você recomendar livros aos clientes, músicas ou filmes, pois eles podem estar interessados tendo em vista suas últimas compras; outra, completamente diferente, é identificar se uma mulher está grávida antes de seus familiares mais próximos tomarem conhecimento. Embora essa

prática seja ilegal dentro das leis atuais dos EUA, críticos expressam preocupações que os dados passem a ser utilizados de formas escusas. Sobretudo porque este tipo de atividade não é tão bem especificado em outras partes do mundo como nos EUA, e pode conseguir cruzar a linha da ficção à realidade, em circunstâncias favoráveis (TENE e POLONETSKY, 2013).

A análise preditiva apresenta ainda o risco de perpetuar as desigualdades e os preconceitos. Os ricos e educados poderão ter o caminho mais rápido; os pobres e desfavorecidos terão ainda mais dificuldades do que antes. A análise preditiva pode acentuar a estratificação social (TENE e POLONETSKY, 2013).

Enquanto os entusiastas de dados veem grande potencial para o uso de *Big Data*, os defensores da ética estão preocupados na medida em que mais e mais dados são coletados sobre as pessoas - tanto consciente ou inconscientemente - pelos proprietários dos dados. A preocupação não é só com o estabelecimento de perfis, mas também com as consequências disso, pois os algoritmos podem tirar as conclusões erradas sobre quem alguém é, como ele/ela pode se comportar no futuro, e como aplicar as correlações que vão surgir na análise dos dados (YAN, 2013).

Uma preocupação adicional levantada pelo *Big Data* é que ela inclina uma escala já desigual em favor de organizações e contra indivíduos. Os grandes benefícios da tecnologia favorecem o governo e as grandes corporações, não os indivíduos (TENE e POLONETSKY, 2013).

### **3.5.2. Privacidade**

Privacidade e segurança cibernética estão entre as principais preocupações para aplicativos de *Big Data*. Privacidade refere-se principalmente a informações pessoalmente identificáveis (PII), isto é, informações que podem ser usadas para identificar um indivíduo. Na Internet, a privacidade dos usuários pode ser dividido em três preocupações: (a) quais informações pessoais podem ser compartilhadas e com quem, (b) se as mensagens podem ser trocadas sem que ninguém mais vê-los, e (c) “se” e “como” alguém pode enviar mensagens de forma anônima (YAN, 2013). A primeira preocupação já foi discutida no tópico de ética, mas as duas seguintes também são importantes, pois abrangem outros pontos a respeito da privacidade.

A questão do anonimato, por exemplo, é uma que tem suscitado muitas discussões na academia. Minelli, Chambers e Dhiraj (2013) acredita que os dados recolhidos para um

fim específico podem ser anonimados e depois ser utilizados para outros fins como a identificação de padrões coletivos, sobretudo quando não há necessidade de voltar aos dados individuais. O desafio, segundo o autor, é que quanto mais anônimos forem os dados, menor utilidade eles têm.

Por outro lado, em 2006, um grupo de pesquisa com sede em Harvard começou a reunir os perfis de 1.700 usuários colegiais do Facebook para estudar como os seus interesses e amizades mudaram ao longo do tempo. Estes dados supostamente anônimos foram lançados para o mundo, permitindo que outros pesquisadores explorassem e analisassem. O que outros pesquisadores rapidamente descobriram foi que era possível tirar do anonimato partes do conjunto de dados, comprometendo a privacidade dos alunos, nenhum dos quais estavam cientes que seus dados estavam sendo coletados (BOYD e CRAWFORD, 2012).

Uma vez que os dados são associados a um indivíduo identificado, eles se tornam difíceis de desassociar. Isso foi demonstrado pelos pesquisadores Arvind Narayanan e Vitaly Shmatikov da Universidade do Texas, que reassociaram recomendações de filmes da Netflix não identificadas com indivíduos identificados, por meio do cruzamento de uma base de dados não identificada com recursos públicos acessáveis na rede (TENE e POLONETSKY, 2013).

Além da questão do anonimato proposta por Yan (2013), Minelli, Chambers e Dhiraj (2013), baseados nos índices culturais desenvolvidos por G. J. Hofstede, adicionam o impacto cultural sobre a noção de privacidade, trazendo ainda maior complexidade ao desafio da privacidade. A aplicabilidade desses índices à privacidade foi validada em um estudo de 2010 da *Carnegie Mellon University*, onde mesmo controlando os resultados para diferenças de idade e gênero, as diferenças entre as culturas foram significativas. Os autores do estudo, Caroline Sheedy e Ponnurangam Kumaraguru, notaram ainda que a noção de privacidade não varia somente com relação à geolocalização, mas também com relação à relevância de tais informações a indivíduos específicos e os seus papéis dentro dos contextos em que a privacidade é discutida. Por exemplo, a sensação de privacidade e a noção de onde traçar fronteiras se alteram em situações em que cada membro da família tem um ou mais dispositivos móveis pessoais. Estes são, portanto, fatores contextuais que afetam a percepção de privacidade. Os contextos estão mudando constantemente entre locais e globais. Indivíduos adotam posturas diferentes de confiança e vontade de compartilhar dependendo desses contextos. O

que compartilhamos em um contexto pode ser claramente não compartilhado em outro. Acadêmicos de Dublin, na Irlanda, para Pittsburgh, Pensilvânia, têm demonstrado a importância do contexto referenciando o conceito de integridade contextual (MINELLI, CHAMBERS e DHIRAJ, 2013).

Tendo em vista esses desafios, a maioria das empresas desenvolveram as suas próprias políticas de privacidade como uma questão de estabelecer um mínimo de "confiança" com os consumidores. Existem diversas variações de sete princípios delineados no documento "Princípios de Porto Seguro entre UE-EUA", que a maioria das empresas já enraizou em sua autorregulação para garantir a privacidade dos dados. A autogovernança, no entanto, pode mudar em breve. Em fevereiro de 2012, a Casa Branca anunciou uma iniciativa para garantir os direitos do consumidor à privacidade<sup>13</sup>. A adesão ao ato ainda é um ato voluntário, não de legislação, mas isso pode mudar à medida que a privacidade se torna um assunto cada vez mais discutido (MINELLI, CHAMBERS e DHIRAJ, 2013).

Além disso, vale lembrar que uma organização pode se esforçar em atuar na questão da privacidade dos dados tanto quanto os dados permaneçam sob seu controle. As organizações que usufruem e dependem de *Big Data* – e efetivamente o tornam ainda maior – enfrentam ainda outro problema: saber o que seus parceiros comerciais farão com os dados que são compartilhados:

Scott Horwitz, diretor sênior de seguros na empresa de crédito de pontuação FICO, aborda a questão da privacidade sob outro ponto de vista, iluminando dois pontos-chave na discussão. Primeiro, padrões emergem quando olhamos para os dados em massa. Os padrões escapariam se as empresas fossem de fato olhar para cada caso individualmente e sem contexto. Este é o preceito fundamental de *Big Data*. Em segundo lugar, *Big Data* não é apenas sobre o aumento de receita ou sobre encontrar insights do cliente, é também sobre identificar fraudes e os responsáveis por ela.

A realidade é que, embora haja tentativas modernas de negócios para tentar controlar toda a experiência do usuário, garantindo a privacidade dos dados do início ao fim, este controle não é possível. Mas é possível influenciar os resultados. A melhor maneira de lidar com questões de sensibilidade é simplesmente permitir que os indivíduos controlem o que "funciona para eles" (MINELLI, CHAMBERS e DHIRAJ, 2013).

---

<sup>13</sup> Em inglês, "*Consumer Privacy Bill of Rights*".

Isso fica claro quando se considera que as pessoas usam os produtos e serviços da Amazon, Apple, Facebook, Google e outras companhias porque querem, não porque existe uma necessidade de uso. Esta é uma distinção fundamental. Talvez a melhor maneira de abordar a privacidade para as organizações, seja criar mecanismos de precaução e, ao mesmo tempo, buscar enxergar as questões sobre a ótica da ética:

*Sempre que se refere aos negócios, eu acho que é fácil violar a privacidade que talvez um melhor enquadramento para a conversa é o da ética. A privacidade dos dados é a coisa que você faz para evitar ser processado, a ética de dados é a coisa que você faz para tornar o seu relacionamento positivos com seus clientes (James Stogdill, O'Reilly Radar, 2005 APUD MINELLI, CHAMBERS e DHIRAJ, 2013).*

A privacidade não é o único problema associado com *Big Data*. A segurança é outro. Grandes conjuntos de dados sensíveis atraem grande interesse por parte de hackers, especialmente quando os dados são tão completos e pessoais. A Apple tem relatado 400 milhões de cartões de crédito do cliente no servidor e a Amazon parece estar no mesmo patamar.

Segurança Cibernética é o corpo de tecnologias, processos e práticas destinadas a proteger redes, computadores, programas e dados de ataque, danos ou acesso não autorizado. Um dos elementos mais problemáticos da segurança cibernética é a natureza rápida e a constante evolução dos riscos de segurança (ROUSE, 2010), devido à constante mudança tecnologias e diversidade de análise de dados de gerenciamento de dados (YAN, 2013).

Além dos riscos acima, as agências federais estão enfrentando outros obstáculos, como as políticas federais sobre privacidade e segurança cibernética que os impedem de beneficiar de grandes potencialidades de dados. Apesar de *Big Data* ter mudado a forma como o governo reúne informações, particularmente a capacidade de coletar grandes quantidades de dados não estruturados, as políticas federais não se adaptaram à mudança. Diretrizes de segurança, políticas de proteção de dados e padrões de compartilhamento de informações variam quase tão amplamente como as missões da agência fazer. Muitas agências estão reagindo à onda de *Big Data* com as políticas obsoletas em vez de tomar uma posição proativa e ajustar suas políticas para atender as necessidades da grande revolução dos dados e abraçar ganhos financeiros potencialmente enormes e insights (KONKEL, 2013 APUD YAN, 2013).

*Big Data* põe grandes riscos à privacidade e à segurança. A coleta de grandes conjuntos de dados pessoais e à utilização do estado da arte de análise implícita crescente preocupação com a privacidade, ao mesmo tempo que aglutinação de informações em determinadas bases de dados aumento o prêmio para os indivíduos com intuítos escusos, e capacidade e interesse em fazer atividade ilícitas. Proteger a privacidade vai se tornar mais difícil conforme a informação é multiplicada e compartilhada cada vez mais por um número cada vez maior de pessoas ao redor do mundo, e proteger os grandes conjuntos de informações oriundos da análise de dados vai ser tornar cada vez mais vital para garantir a segurança. Por causa disso, esses são importantes candidatos à fatores relevantes na adoção e uso da tecnologia.

## 4. MODA: MODELO NO QUAL FOI BASEADO O PENSA

*In God we trust, everyone else bring data (Michael Bloomberg).*

### 4.1. ANTECEDENTES

O desenvolvimento do uso de *Big Data* na gestão do governo da cidade de Nova York teve três importantes momentos: o desenvolvimento do modelo Compstat no início dos anos 90, a propagação do modelo usado para outras agências no início da década de 2000, e o desenvolvimento do modelo que ficou conhecido como “Gov. 2.0”, no mandato do prefeito Michael Bloomberg.

#### 4.1.1. O Desenvolvimento do Modelo Compstat

Durante a década de 80, o Departamento de Polícia de Nova York (NYPD) enfrentou diversas limitações de recursos, que ocasionaram, até o início dos anos 90, numa redução da força policial em 34% e um aumento da criminalidade sem precedentes. Em 1993, Rudolph Giuliani foi eleito prefeito da cidade com uma proposta de redução da criminalidade. Uma das primeiras medidas do novo prefeito foi autorizar a contratação de força policial e aumentar as verbas destinadas ao departamento, de modo que foi possível manter os padrões de qualidade do atendimento às emergências anteriormente estabelecidos e a abordagem de novas estratégias policiais (GILSINAN e STEPAN, 2014).

Para levar a cabo a nova fase, foi alçado ao comando da força policial o antigo chefe da força policial de trânsito, William Bratton, que havia experimentado uma redução dos crimes ocorridos no Metrô ao dar maior ênfase à incidência dos crimes de menor relevância – como, por exemplo, vadiagem – por acreditar que a punição severa a esses crimes desestimularia a ocorrência de crimes mais graves (GILSINAN e STEPAN, 2014).

De acordo com Bratton, entre as novas estratégias pretendidas estava a concepção de que, além de dar resposta ao crime, o departamento de polícia devia se antecipar ao crime, adotando medidas preventivas. O grande desafio nesse sentido era a demora da análise dos dados: embora houvesse muitos dados oriundos das ligações para o 9-1-1, número de emergência americano, a polícia só redigia relatórios trimestrais. Por

consequente, embora utilizasse os padrões oriundos da análise de dados para desenvolver estratégias de combate ao crime, essas estratégias eram sempre formuladas em crimes que tinham ocorrido há pelo menos um mês antes, podendo chegar a quatro meses passados (GILSINAN e STEPAN, 2014).

Para superar esse desafio, em abril de 1994 foi desenvolvido e instalado um sistema de informação que atualizava em tempo real as estatísticas de crime em Nova York, por meio de um sistema de contabilidade, o Compstat, que permitia a visualização de onde os padrões de crimes estavam ocorrendo e onde os padrões de prisão estavam acontecendo. Reuniões bissemanais eram feitas para que fosse discutido o que estava sendo realizado para combater a criminalidade com base na informação agora disponível. O Compstat mudou a forma como os dados eram coletados, os recursos eram utilizados e como os comandantes da força policial eram responsabilizados. Representou, portanto, uma mudança de mentalidade – de resposta ao problema para a prevenção do mesmo - mudança esta que iria se propagar por agências similares (GILSINAN e STEPAN, 2014).

#### **4.1.2. Propagação do Modelo Compstat para outras agências**

Uma das primeiras agências a adotar o sistema de análise de dados após o departamento de polícia foi o departamento de Parques e Recreação<sup>14</sup> que, em março de 1997, desenvolveu o seu próprio modelo de coleta e análise de dados. Com o desenvolvimento do novo sistema, o departamento de Parques e Recreação pode desenvolver um sistema de avaliação semanal, além de conseguir dobrar o número de parques inspecionados em pouco tempo. O sucesso foi tamanho que em 2002 o departamento havia expandido sua coleta e análise de dados para outras ocorrências tais como manutenção de veículos, pessoal, alocação de recursos, entre outros (GILSINAN e STEPAN, 2014).

Parques e Recreação foi o primeiro departamento a adotar um sistema baseado no Compstat, mas o Departamento de Bombeiros de Nova York (FDNY) também o fez: apesar de sua reconhecida fama como um dos departamentos de bombeiros de maior sucesso nos EUA, o sistema de inspeção de prédios do FDNY estava falido em 2007. Essa situação ficou clara após um incêndio em um antigo prédio do Deutsche Bank, que estava vazio e previsto para demolir há algum tempo: foram necessárias sete horas de 475 bombeiros para cessar o fogo, devido a uma falha no sistema de direcionamento de

---

<sup>14</sup> Em inglês, Parks and Recreation.



água para cessar os focos de incêndio do prédio, o que poderia ter sido prevenido por uma inspeção regular (GILSINAN e STEPAN, 2014).

A falta de inspeção do prédio que tornou o combate ao incêndio do Deutsche Bank uma situação ainda mais delicada era apenas um dos desafios do FDNY: o departamento contava com apenas 350 inspetores para inspecionarem 300 mil edificações na cidade inteira. Havia, portanto, limitação de recursos. Além disso, o sistema de avaliação era antigo e os critérios estavam inalterados há cerca de 60 anos. Havia falta de dados ou existência de dados muito ruins – muitos dos prédios possuíam apenas aprovações ou desaprovações, sem indicações de causas, justificativas ou desenvolvimento do processo de inspeção – por fim, os dados eram todos físicos, o que tornavam a análise ainda mais difícil (GILSINAN e STEPAN, 2014).

Para superar esses desafios, uma equipe da IBM foi contratada para mapear o *status quo* e digitalizar toda a informação. Foi importante envolver os inspetores desde cedo na reforma dos processos para que o problema do *gap cultural* fosse superado desde cedo. Em meados de 2010, o sistema estava pronto. No entanto, para fazer a avaliação de quais prédios tem mais risco de iniciar um incêndio, ainda havia outro desafio: muitos dos dados que permitiam uma análise de risco da situação real da edificação pertenciam a outras agências municipais. Por exemplo, edificações antigas em bairros pobres e de alta criminalidade apresentavam maior risco: os dados de idade da edificação, no entanto, pertenciam ao Departamento de Edificações (DOB<sup>15</sup>), bem como os dados de criminalidade pertenciam ao NYPD (GILSINAN e STEPAN, 2014).

### **4.1.3. Desenvolvimento do Modelo GOV 2.0**

Enquanto o FDNY enfrentava o desafio de convencer outras agências a liberar seus dados para análise, Michael Bloomberg substituiu Giuliani na prefeitura de Nova York em 2002, com uma prioridade: promover o uso de dados para um governo melhor. Uma das primeiras inovações do novo prefeito foi a instituição do 3-1-1 em 2003, um número que tal qual o 9-1-1 era voltado para denúncia e comunicação da população – neste caso, de ocorrências não emergenciais. O serviço foi ainda mais desenvolvido posteriormente, com um *upload* de um mapa interativo de dados na *web* que indicava quais tipos de problemas estavam acontecendo aonde (GILSINAN e STEPAN, 2014).

---

<sup>15</sup> Em inglês, Buildings Department.

Outra iniciativa foi a transformação do tradicional Relatório de Gestão da Prefeitura, oriundo do Sistema de Planejamento e Relatório de Gestão do Governo (MMPRS), em algo mais amigável ao usuário por meio de uma plataforma simples, intuitiva e online: até então o relatório era visto como um volumoso relatório de estatística que devia ser publicado duas vezes ao ano como um exercício obrigatório de prestação de contas (GILSINAN e STEPAN, 2014).

O desenvolvimento dessas ferramentas e iniciativas tornou mais acessível os serviços públicos e a gestão pública, facilitando para os cidadãos não somente obter os dados, mas também avaliar as políticas e diretrizes da prefeitura. O governo, no entanto, achava necessário se antecipar aos problemas levantados pelos cidadãos (GILSINAN e STEPAN, 2014).

Em fevereiro de 2013, Bloomberg anunciou a criação do Departamento de Análise de Dados da Prefeitura (MODA<sup>16</sup>), uma pequena equipe subsidiada e situada na prefeitura que tinha como objetivo combinar, interrogar e sintetizar os dados de diferentes fontes – tais como diferentes agências municipais – para aumentar a eficiência e a eficácia das operações e serviços governamentais. O intuito era justamente atuar na resolução de problemas que estavam além das responsabilidades individuais de cada um dos departamentos e secretarias, e fazê-lo por meio das informações coletadas pelas agências (GILSINAN e STEPAN, 2014).

A equipe foi inicialmente composta por seis analistas de dados que buscavam maneiras inovadoras de resolver problemas. Os integrantes se baseavam nos *insights* de que os dados relevantes para o desempenho de uma agência poderiam estar estocados em outras agências. De acordo com o líder do projeto, Michael Flowers, o desafio de compartilhamento de dados residia em quatro categoriais – técnica, cultural, política e legal – das quais ele era responsável, sobretudo, pela primeira, tendo em vista que reunir dados de 40 diferentes agências municipais não era uma situação simples, tendo em vista que os dados eram formatados de maneiras diferentes, com periodicidades diferentes, cobrindo períodos distintos, entre muitas outras características (COPELAND, 2014).

O MODA desenvolveu diversos estudos como, por exemplo, estudos que objetivavam identificar fraudes médicas e violações de impostos. Um dos estudos realizados foi para

---

<sup>16</sup> Acrônimo do inglês “Mayor’s Office of Data Analytics”.

o desenvolvimento de um modelo que pudesse analisar as reclamações recebidas via o 3-1-1 e identificar aquelas com maior probabilidade de resultar na identificação de propriedade de maior risco, estudo esse que foi solicitado pelo Departamento de Edificações (DOB). Ao analisar conjuntos de dados diferentes, MODA conseguiu identificar indicadores preditivos de edifícios mais perigosos. Os analistas foram capazes de criar um modelo de previsão de risco que permite aos inspetores do DOB encontrar mais de 70% dos piores edifícios, visando apenas 30% dos edifícios de risco, um resultado impressionante, sobretudo, quando é considerado que apenas cerca de 8% das ligações recebidas por ano no 3-1-1 identificavam apartamentos ilegais de alguma maneira (COPELAND, 2014).

De todos os estudos realizados pelo MODA, um dos que ficou mais famoso foi o de identificação de restaurantes que despejavam ilegalmente óleo nos esgotos. Usando os dados da Comissão de Integridade dos Negócios de uma agência municipal – que tinha contratado empresas para reciclar o óleo e tinha um registro de restaurantes cadastrados para coleta do óleo – e os dados oriundos obtidos nos esgotos como teor de óleo na água, a equipe foi capaz de identificar as principais áreas para inspecionar por restaurantes com práticas ilegais de despejo de óleo nos esgotos, permitindo à cidade economizar milhões de dólares – usando dados de acesso público (COPELAND, 2014).

## **4.2. MODA – MAYOR’S OFFICE OF DATA ANALYTICS**

### **4.2.1. Premissas e Objetivos**

O modelo do MODA foi desenvolvido com base em algumas premissas, destacadas a seguir (COPELAND, 2014):

- Não requerer extensa (e cara) nova tecnologia ou colocação de sensores em cada rua, mas numa melhor utilização dos dados já recolhidos.
- Não envolver mudança fundamental da natureza das atividades realizadas pelo pessoal da linha de frente, mas priorização inteligente de seu trabalho.
- Não insistir na pureza de dados e padrões abertos (formatos e esquemas para a gravação de dados comuns), mas na abrangência e na completeza dos dados.
- Não apostar em uma nova "cidade inteligente" e em modelo de negócios radicais, mas em testes e escala ideias que fornecem retorno comprovado sobre o investimento (ROI).

- Dependem menos da especialização tecnológica e muito mais em uma forte liderança política das figuras de maior proeminência na cidade e do governo local.

Como pode ser visto, o MODA foi estabelecido então com o intuito de contribuir para a melhora da cidade por meio da análise de dados. Para tal, eles objetivavam influenciar tão pouco quanto possível como as atividades eram realizadas, focando apenas na melhor utilização de dados já coletados e na priorização dos trabalhos das agências municipais. Em outras palavras, o esforço não era no sentido desenvolver uma *smart city*, e sim focar em atividade que fornecessem retorno comprovado. Os objetivos específicos do MODA são especificados à seguir (COPELAND, 2014):

1. Ajudar as agências municipais a desenvolver melhor os seus serviços.
2. Compartilhar dados com as agências municipais e encorajar o uso das melhores práticas de análise de dados.
3. Atuar como corretor de dados, tendo em vista que tem acesso a inúmeros dados de diversos sistemas legados e pode coletar e tratar os dados para transmitir para outras agências.
4. Utilizar a análise de dados para entregar *insights* a respeito do desenvolvimento econômico – isto é, não limitar-se à solução dos problemas atuais das agências.
5. Modelar o impacto de uma nova proposta de lei.
6. Ajudar na resposta à desastres e na recuperação necessária.
7. Prover acesso aos dados.

Destes objetivos, àquele que trouxe maior reconhecimento para o MODA devido ao impacto de suas atividades foi justamente o primeiro, devido ao valor que ele agregou às equipes prestadoras de serviço à sociedade, tais como bombeiros e agentes de trânsito, ao levar a análise de dados para suas atividades cotidianas com o intuito de melhorá-las. Para melhor desempenhar este objetivo, o MODA desenvolveu um modelo operacional de 10 passos operacionais, explicado à seguir (COPELAND, 2014):

- (1) Entender o funcionamento das atividades diárias do órgão. A equipe do MODA desempenha um trabalho de sombra às equipes dos prestadores de serviços quando vai desenvolver um estudo com/para eles de modo a entender a natureza dos serviços prestados, como os recursos são alocados e utilizados, quais fatores

e como eles influenciam na priorização do serviço e como os dados são gravados nos sistemas de informação dos prestadores de serviço.

- (2) Identificar áreas em que os dados podem ajudar e então verificar quais dados podem ajudar com a melhora proposta.
- (3) Criar um plano de projeto de modo a encontrar uma concordância com o prestador de serviço.
- (4) Entender o contexto de dados.
- (5) Estabelecer um acordo formal com o prestador de serviço, onde é explicado o propósito do compartilhamento de dados, os tratamentos realizados nos dados para garantir a privacidade e segurança dos dados. O documento também serve para trazer transparência e comprometimento das duas equipes envolvidas com o projeto.
- (6) Integrar os dados necessários para a análise.
- (7) Testar hipóteses.
- (8) Revisão e validação do estudo junto ao prestador de serviços.
- (9) Automatização do processo.
- (10) Implementação da solução e delegação da responsabilidade pelo modelo de dados ao prestador de serviços.

Um modelo de processos de como esses dez passos eram efetivamente executados ao longo de um estudo do MODA segue no anexo VI.

#### **4.2.2. Equipe e Estrutura**

Para realização do modelo operacional e o alcance dos objetivos previamente citados nesse documento, o MODA conta atualmente com uma pequena equipe de nove pessoas, incluindo analistas e uma equipe de suporte técnico e administrativo. Os analistas têm um mix de conhecimentos técnicos, econômicos e de ciência da computação. Em 2013, os papéis da equipe incluíam um Diretor-chefe de Análise de Dados e um Diretor-chefe de Plataformas Abertas; um Vice-diretor; um Analista-chefe; um Programador-chefe; um assessor especial do Diretor-chefe de Análise de dados; um conselheiro sênior do Diretor-chefe de Análise de dados (COPELAND, 2014).

É importante, no entanto, destacar algumas características da equipe, tais como o fato de a equipe ser pequena, transversal, ligada direta e somente à prefeitura, e não ser direcionada à solução dos problemas das diversas agências municipais. Pelo contrário, o

MODA foi idealizado para reverter algo que comumente tem sido observado: a transformação da compartimentação dos serviços em agências municipais (feito originalmente para garantir eficiência e eficácia na prestação dos serviços) em silos que impeçam a troca de informações, dados e *insights*. Por conseguinte, o MODA foi formatado de modo a auxiliar as secretarias fornecendo *insights* externos a elas – que muitas vezes podem estar em outras secretarias –, e não para resolver os problemas que elas enfrentam (COPELAND, 2014).

Longe de criar uma camada adicional significativa da burocracia, o modelo New York MODA é enxuta e altamente eficiente (COPELAND, 2014).

### **LIÇÕES DO MODA PARA FUTURAS EQUIPES DE ANÁLISE DE DADOS NO SETOR PÚBLICO**

Por fim, algumas das lições adquiridas com os esforços de estabelecimento e implementação do MODA e uso de *Big Data* realçadas pelos participantes do MODA são (COPELAND, 2014):

- (1) Um forte suporte da alta-gestão é essencial;
- (2) Os modelos devem ser formatados pela experiência e expertise dos atores que executam a tarefa;
- (3) É importante focar em projetos cujos resultados provêm Retorno sobre Investimento;
- (4) Deve-se começar por projetos pequenos e com medidas e mensurações que todos podem entender e apoiar;
- (5) Não se deve modificar o trabalho dos prestadores de serviços;
- (6) Usar dados não requer um grande número de pessoas especializadas ou novas camadas de burocracia;
- (7) Usar dados não requer a aquisição de tecnologia de ponta;
- (8) Qualquer organização que requer acesso aos dados do MODA deve primeiro compartilhar seus próprios dados;
- (9) Todos os dados devem ser geolocalizados;
- (10) Nenhuma parte da extração de dados ou do processamento de dados deve requerer a ação humana.

## 5. MÉTODO

### 5.1. TIPO DE PESQUISA

A pesquisa se iniciou com uma revisão bibliográfica sobre o tema, abrangendo a literatura acadêmica de outros sistemas de informação, sobretudo sistemas de BI. A partir dessa literatura, foi construído o conceito de *Big Data* adotado nesta pesquisa, bem como foi remontada a evolução conceitual e histórica do grupo de tecnologias analíticas consideradas parte do *Big Data*. Em seguida foi composta uma lista de potenciais fatores relevantes na adoção e uso das tecnologias, sem a pretensão de ser uma lista exaustiva tendo em vista o caráter exploratório do estudo, devido ao pouco conhecimento acumulado sobre o assunto. Em seguida foi realizado um estudo de caso para investigar os fatores relevantes na adoção e uso de *Big Data*.

O método escolhido foi o do estudo de caso, por ser recomendado a estudos exploratórios e em fase inicial de investigação (YIN, 1989). O estudo de caso se caracteriza como um tipo de pesquisa cujo objeto é uma unidade que se analisa profundamente (GODOY, 1995). Assim, esse modelo tem se tornado uma estratégia utilizada quando os pesquisadores procuram responder "como" e "por que" os fenômenos ocorrem, quando o investigador tem pouco controle sobre os eventos, e quando o foco da pesquisa é um fenômeno inserido em algum contexto da vida real (YIN, 1989) – características bem próximas da presente pesquisa.

Em específico, a utilização de um único caso é apropriada em algumas circunstâncias: quando se utiliza o caso para determinar se as proposições de uma teoria são corretas; quando o caso sob estudo é raro ou extremo, ou seja, não existem muitas situações semelhantes para que sejam feitos outros estudos comparativos; quando o caso é revelador, ou seja, quando ele permite o acesso a informações não facilmente disponíveis (STAKE, 2000 APUD ZANNI, MORAES e MARIOTTO, 2011). O estudo de caso desenvolvido neste trabalho é um estudo de caso único, tendo vista que as implantações de *Big Data* no Brasil ainda não são numerosas e que sua realização pode permitir o levantamento de proposições a serem testadas em pesquisas futuras.

## 5.2. O CASO SELECIONADO

O caso estudado no presente trabalho foi a implementação e uso de *Big Data* no Pensa – um departamento ligado à Casa Civil da Prefeitura do Rio de Janeiro, onde se deu o primeiro caso de adoção do *Big Data* pelo poder público no Brasil com o intuito de melhora dos serviços providos e criação e desenvolvimento de uma visão estratégica de longo prazo. A prefeitura da cidade (PCRJ) administra o município que é a maior rota de turismo internacional de turismo do Brasil. Com aproximadamente 6,5 milhões de habitantes e uma área geográfica de 1,2 milhões de metros quadrados, o Rio de Janeiro é a segunda maior metrópole do país, a sexta maior da América Latina e a trigésima quinta do mundo. Sua administração municipal reúne 29 secretarias, 8 empresas públicas, 5 autarquias e 6 fundações.

A adoção do *Big Data*, no caso em questão, ocorreu em uma estrutura à parte do organograma da prefeitura para que fosse patrocinada diretamente pela Casa Civil, de maneira que pudesse apresentar resultados operacionais tão rápido quanto possível, sem perder de vista a visão e o desenvolvimento estratégico. Os resultados mencionados são os estudos realizados junto a outros departamentos e secretarias da prefeitura.

## 5.3. COLETA E ANÁLISE DE DADOS

O estudo de caso foi então desenhado para ser realizado por meio de entrevistas, uma das principais fontes de informação nesses casos (YIN, 1989). O roteiro desenvolvido não buscou apresentar uma lista exaustiva dos fatores relevantes na adoção e no uso de *Big Data*, devido ao caráter exploratório do estudo. O roteiro está apresentado no anexo I.

A equipe do PENSA conta atualmente com cinco funcionários, um dos quais é o gestor da equipe. Como a equipe já teve 7 funcionários e existem pelo-menos 4 ex-funcionários, objetivou-se a princípio entrevistar todos os funcionários e ex-funcionários para ter uma visão mais ampla da dinâmica de funcionamento do PENSA desde o momento da adoção do *Big Data* até o presente momento. Devido a difícil acessibilidade a alguns funcionários e ex-funcionários, no entanto, isso não foi possível, de modo que o objetivo passou a ser entrevistar o maior número possível de



funcionários, ex-funcionários e funcionários de áreas clientes do departamento para ter total compreensão dos aspectos relevantes à adoção e uso de *Big Data*.

Foram realizadas um total de 10 entrevistas com os atuais integrantes do departamento estudado, ex-integrantes da equipe e funcionários de áreas correlatas ao departamento – um funcionário da Casa Civil e três funcionários pertencentes às duas principais áreas clientes do departamento (a SMTR e o 1746). Para garantir a fidedignidade das informações e a privacidade dos entrevistados, os funcionários foram categorizados em funcionários do departamento (grupo no qual consta além dos funcionários propriamente ditos, ex-funcionários e o gestor do PENSA), funcionários de áreas clientes (funcionários da SMTR e do 1746) e funcionário da Casa Civil (representante da alta gestão), conforme a tabela a seguir.

Funcionário E1	Funcionário ou ex-funcionário do PENSA
Funcionário E2	Funcionário ou ex-funcionário do PENSA
Funcionário E3	Funcionário ou ex-funcionário do PENSA
Funcionário E4	Funcionário ou ex-funcionário do PENSA
Funcionário E5	Funcionário ou ex-funcionário do PENSA
Funcionário AC1	Funcionário de áreas clientes
Funcionário AC2	Funcionário de áreas clientes
Funcionário AC3	Funcionário de áreas clientes
Representante da Casa Civil	Representante da Casa Civil

**Tabela 5:** Tabela de Funcionários do PENSA e das Áreas Clientes entrevistados (**Fonte: autor**)

As entrevistas foram gravadas com consentimento do entrevistado e foi dito a ele que sua identidade seria mantida em sigilo. A quantidade de entrevistas realizadas seguiram o critério de saturação, que designa o momento em que o acréscimo de dados e informações em uma pesquisa não altera a compreensão do fenômeno estudado. Ela pertence à esfera da validação objetiva - adequação de uma conjectura a uma explicação lógica (CRESWELL, 1998; GUEST, BUNCE e JOHNSON, 2006; THIRY-CHERQUES, 2008a).

Além das entrevistas, foram consultados materiais secundários disponibilizados pelos entrevistados tais como os decretos de criação do PENSA e de disponibilização de dados para o PENSA e materiais disponibilizados na imprensa tais como entrevistas para a revista *Veja Rio*, jornais e blogs. Foram consultados também materiais que estão disponíveis na internet sobre o MODA – departamento de análise de dados da prefeitura de Nova York que serviu de modelo para o planejamento e criação do PENSA –, tais como decreto de criação e modelo de processos no funcionamento.

Vale destacar ainda que as entrevistas foram transcritas e totalizaram mais de 200 páginas. O material coletado está à disposição para consulta mediante contato com o autor do trabalho. Os dados coletados foram analisados pela técnica de análise de conteúdo. As categorias apresentadas foram definidas de acordo com a revisão bibliográfica e depois complementadas de acordo com o surgimento de novas informações a partir das entrevistas.

#### **5.4. PENSA – SALA DE IDEIAS**

Baseada na New York City Mayor's Office of Data Analytics (MODA) – iniciativa de *Big Data* e análise de dados da cidade de Nova York –, a Prefeitura de Janeiro idealizou uma área de *Big Data*, o “Pensa – Sala de ideias”, conforme pode ser visto nas entrevistas realizadas.

O PENSA foi criado em maio de 2013 e pelo que eu sei surgiu da cabeça do prefeito e do Pedro Paulo que na época era secretário da Casa Civil. A gente tem 1746 que é bastante inspirado no 311 que é o atendimento ao cidadão lá de Nova Iorque. Temos tentativas também de importar um pouco outros modelos de parcerias público-privadas também e uma das coisas que chamou atenção foi o MODA, que é o PENSA lá de Nova Iorque. A prefeitura tem um banco de dados e eles muitas vezes são centralizados para uso das próprias secretarias, mas olhando para a cidade de Nova Iorque eles quiseram criar uma equipe que pudesse centralizar e usar esses dados para propor políticas públicas inovadoras baseadas em dados (Ex-funcionário do PENSA).

Criado pela secretaria de Casa Civil da Prefeitura da Cidade do Rio de Janeiro por meio do Decreto Nº 37.215 de 3 de junho de 2013, o PENSA é um grupo ligado à Secretária da Casa Civil do Rio de Janeiro responsável por analisar informações geradas por toda a população da cidade do Rio de Janeiro e pelos mais variados sensores com o intuito de desenhar políticas e soluções para a cidade. Entram aí índices obtidos pelas diversas secretarias e órgãos públicos, somados aos que chegam via câmeras de controle, dispositivos de GPS, radares e canais como a central telefônica e *mobile* de atendimento ao público 1746, além das redes sociais - só para citar algumas das fontes que geram mais de um milhão de dados da cidade por dia (VEJA RIO, 2014)

De acordo com o decreto supracitado, e que segue no anexo VII, compete à equipe PENSA – Sala de ideias:

I - Entender os serviços prestados pela Prefeitura bem como seus desafios;

II - Trabalhar com a ampla base de dados da Prefeitura, estruturados ou não, oriundos dos diversos órgãos municipais;

III - Dar suporte de análise de dados aos times dos diferentes órgãos na busca de novas soluções;

IV – Apontar possíveis soluções aos gestores da prefeitura para problemas;

V - Definir e implementar planos de ação, estruturais e tecnológicos, para ser reconhecido como área de vanguarda em *Big Data*;

VI - Inovar nas soluções propostas;

VII - Criar convênios com universidades, instituições e empresas quando se fizer necessário e que tragam conhecimento e inovação para os processos em questão;

VIII - Formar profissionais da área de *Big Data*.

A Prefeitura do Rio de Janeiro passou a apostar no *Big Data* para conhecer melhor as demandas de sua população e auxiliar o poder executivo a tomar as melhores decisões para o município (SOUZA, 2014). Para tal foi formatada uma equipe multidisciplinar – à semelhança da iniciativa do MODA – pequena, focada em resultados, estruturada de maneira centralizada e subordinada diretamente à Secretaria da Casa Civil.

No que tange a estrutura, foi escolhida uma centralizada, em detrimento das estruturas descentralizadas e federadas, devido aos objetivos do PENSA. Tal como o MODA, O PENSA não objetiva resolver os problemas já enfrentados pelas secretarias ou auxiliar no desenvolvimento de suas atividades diárias – que favoreceria uma estrutura descentralizada. Não é tampouco criar uma estrutura paralela que crie *enforcement* suficiente para que as secretarias sejam obrigadas a usar dados e prestar contas de como e porque tomaram determinadas decisões – o que favoreceria o uso de uma estrutura federada. O objetivo do PENSA é servir como um departamento transversal que aumente a comunicação entre as secretarias e, tal qual o MODA, reverta a transformação da compartimentação das secretarias em silos que impeçam a troca de informações, dados e *insights*. Por conseguinte, o PENSA foi formatado em estrutura centralizada de modo a auxiliar as secretarias fornecendo *insights* externos a elas – que muitas vezes podem estar em outras secretarias –, e não para resolver os problemas que elas enfrentam.

Além disso, é importante citar que o PENSA se concentrou no uso de tecnologias *Open Source*. Essa opção está associada à diminuição dos custos de infraestrutura, que são

pequenos e limitam-se apenas à manutenção do servidor robusto o suficiente para a realização das análises, conforme pode ser observado nos trechos destacados a seguir.

De início, nós fizemos todo um planejamento das licenças, aí foi ver o custo de cada um, muito elevado então vamos para essa outra opção que é uma opção muito boa (Funcionário E1 do PENSA).

Custo é um critério forte e o outro é por conhecimento nosso (Funcionário E2).

A missão do Pensa é aprimorar a prestação de serviços ao cidadão através de soluções inovadoras fruto de pesquisa, análise e criação de correlações de dados internos e externos, levando em consideração as demandas dos diversos órgãos e secretarias (MATHEUS, 2013). E o grupo faz isso sem envolver-se diretamente com a atividade operacional das diversas secretarias da cidade, e sim por meio da análise de dados e produção de estudos, conforme esclareceu o coordenador do grupo em entrevista à revista Veja:

"O Pensa não executa projetos. Ele propõe e testa hipóteses, avalia eventuais ganhos e desenvolve, em conjunto com os demais órgãos da prefeitura, ideias que possam ser aplicadas em curto prazo e que tragam benefícios ao maior número possível de cidadãos" (CERDEIRA, Pablo em entrevista à Revista Veja).

Ao longo do primeiro ano de atuação, o grupo apresentou mais de 15 projetos. O primeiro objeto de estudo do grupo foi o mapeamento das árvores da cidade com o intuito de tentar antever quais as árvores tinham maior probabilidade de cair e, assim sendo, criar uma ação preventiva. A partir da informação de que as árvores ficavam vulneráveis a partir de ventos superiores a 40km/h, passou-se a emitir um sistema de alerta para a Comlurb toda vez que chega-se nesse índice. O órgão também passou a conhecer as regiões da cidade em que há maior número de registros de queda e, de forma preventiva, intensificou o serviço de poda nos locais mais comumente afetados.

Entre outros estudos destacam-se ainda estudos na área de transportes, na área de saúde, na área de educação, na área de conservação e na área de limpeza urbana, tais como estudos de combate à dengue, alagamentos, estacionamentos, rotas de ônibus e disposição dos pontos de ônibus e estudos sobre a movimentação da população durante os Réveillons (VEJA RIO, 2014).

Entre os resultados desses diversos estudos, destaca-se o do estudo sobre o combate à dengue, que foi encaminhado aos órgãos de limpeza, saúde e educação, e deu origem a várias ações práticas de prevenção e de conscientização da população que culminaram na redução em 94% dos números de casos de dengue na cidade (VEJA RIO, 2014).

Destaca-se ainda a instalação de 20 relógios digitais com dicas de rotas alternativas para fugir dos costumeiros congestionamentos da cidade. Os relógios indicam informações em pontos com grande fluxo de veículos a partir de informações geradas pelos dados do Waze, sendo que o cálculo do tempo de percurso é processado pela ferramenta do Google, que leva em conta dados do trânsito monitorados em tempo real (ANDRADE, 2014).

## 6. ANÁLISE DE DADOS

### 6.1. FATORES ORGANIZACIONAIS

#### 6.1.1. Alinhamento à Estratégia

Como pode ser visto na revisão bibliográfica, benefícios e oportunidades gerados pelo *Big Data* podem ser encontrados em todos os departamentos de qualquer organização, ao longo de todos os setores e indústrias da economia. De acordo com Bughin, Livingston e Marwaha (2011), isso não significa que as empresas devam aderir de maneira intempestiva ao *Big Data*. É necessário que a estratégia de uso do *Big Data* esteja intimamente conectada a estratégia da empresa, e para isso é necessário um extenso planejamento para decidir como os recursos analíticos devem ser aplicados para obter o resultado esperado (NOVO e NEVES, 2013).

Nesse sentido, pôde ser verificado, por meio das entrevistas, que a criação do PENSA e a adoção do *Big Data* como solução tecnológica estão conectadas à estratégia da Prefeitura de gestão por resultado por meio de uma atuação transversal, centralizada e distante do conceito tradicional de secretarias praticamente independentes uma das outras, conforme pode ser visto nos trechos destacados a seguir.

(Existe) um esforço muito grande para se trabalhar com planejamento estratégico dentro do governo. Foi feito o primeiro planejamento estratégico em 2009, onde se estabeleceram uma série de metas e projetos para o governo, [...] no início de 2012 a gente lança uma extensão do plano, de 2013 a 2016 com uma visão até 2030 e isso sinaliza muito claramente, do ponto de vista da gestão, a estratégia do governo era uma estratégia que ia privilegiar a gestão através de resultado, a gestão através de metas [...] no fundo, quando você tem a criação do PENSA, ele está totalmente relacionado a esse conceito de uma gestão por resultado (Representante da Casa Civil).

Desde que o prefeito assumiu o 1º mandato, ele vem trabalhando uma série de estratégias que quebram um pouco aquele conceito tradicional de gestão da prefeitura, em que você tem um prefeito e uma série de secretarias praticamente independentes [...] Esse trabalho de puxar pra junto da administração central a organização de projetos, faz com que você tenha que ter também um órgão como é o PENSA que trabalhe também transversalmente em todos os órgãos para poder ter os insights, ter os insumos que a Casa Civil e o gabinete do prefeito precisam para que os projetos sejam articulados (Funcionário E1 do PENSA).

A importância do alinhamento da implementação da tecnologia com a estratégia da organização ficou ainda mais clara pela entrevista de outro funcionário, que deixou mostrar que o PENSA é como uma ferramenta de auxílio à transformação do Rio de

Janeiro na visão de ser a melhor cidade do hemisfério sul para trabalhar e morar até 2030.

As metas do PENSA são ligadas a metas estratégicas de diversos valores, mas assim em relação á visão da prefeitura que é tornar (o Rio de Janeiro) a melhor (cidade) para trabalhar e morar no hemisfério sul até 2030 (...) o PENSA é como uma ferramenta, como uma equipe que pode ajudar a transformar a cidade para atingir esse objetivo (Funcionário E5 do PENSA).

Essa visão do PENSA como uma ferramenta também está em concordância com o objetivo estratégico da prefeitura de tornar o Rio de Janeiro a melhor cidade do hemisfério sul, pois o PENSA pode reforçar o fluxo de informação na sociedade e fazer uso disso para identificar as melhores medidas a serem tomadas pela prefeitura.

Isso sempre foi uma percepção minha. A cidade já gerava muitos dados, talvez não estivéssemos usando elas de maneira inteligente [...] É importante ter fluxo de informação para o sistema complexo conseguir se auto-organizar de uma ótima maneira. Eu senti que esse fluxo de informação nas cidades, no governo, era muito fraco, deveria existir alguma maneira de reforçar esse fluxo de informação. O PENSA vem bem de encontro (a essa necessidade) (Funcionário E3 do PENSA).

Esses achados corroboram com a visão de que o alinhamento do sistema de *Big Data* com a estratégia da empresa é um fator relevante na sua implementação e uso, pois, conforme descreveu Yeoh & Koronius (2010), deixam de ser cometidas as principais falhas na implementação de sistemas integrados que são a falta de alinhamento com a visão de negócios da empresa e o fracasso em atender os objetivos principais da implementação.

O conteúdo das entrevistas começou então a revelar algumas características interessantes a respeito da iniciativa por trás do PENSA: se por um lado, havia um uso estratégico do novo departamento e o *Big Data* era visto como uma ferramenta a mais para a realização da visão de longo-prazo da prefeitura, por outro, não existia a comunicação clara um objetivo específico, mensurável e verificável – um dos principais benefícios do alinhamento estratégico justamente por nortear e sustentar o compromisso de todos os envolvidos com a implementação de acordo com a revisão bibliográfica. Isso pode ser verificado pela contradição nas respostas de dois funcionários do PENSA quando questionados sobre o objetivo do PENSA. Enquanto um funcionário afirma que a prefeitura tem um plano de metas e o PENSA não faz parte dele, o outro pontua um objetivo bem genérico e não mensurável.

Tem um plano de metas e a gente não está à parte disso (Funcionário E4 do PENSA).

O uso desses dados. Então você fazendo uso desses dados e transformando esses dados em informação, que tipo de informação você iria trabalhar? Uma informação que tivesse aquela finalidade de melhorar a cidade. Então o nosso objetivo aqui o tempo todo é analisar dados, transformar em informações para o cidadão, melhorar a vida do cidadão (Funcionário E3 do PENSA).

Essa falta de clareza na comunicação do objetivo do PENSA fica ainda mais evidente quando são comparados os trechos transcritos acima com a resposta dada pelo representante da Casa Civil – secretaria à qual o PENSA está subordinado –, que afirmou um objetivo específico diferente daquele dito pelos funcionários da equipe: o apoio de fatos e dados para a tomada de decisão – afastando-se de achismos, como ele mesmo definiu.

Uma gestão que sabe exatamente onde quer chegar e uma gestão que privilegia fatos e dados para a sua tomada de decisão, menos preocupados com achismos, com reportagem, estão mais preocupados com informação inteligente e sofisticada para você tomar a decisão.

Nesse sentido, a entrevista com o funcionário E5 foi fundamental para compreender que o objetivo específico foi estabelecido, embora não seja claro ou não seja bem comunicado. Ao mesmo tempo em que afirmou que o decreto de criação do PENSA como órgão da prefeitura é claro sobre o objetivo – uso de dados e análise de dados –, quando questionado sobre a existência de possíveis conflitos com outros órgãos e secretarias da prefeitura, o funcionário em questão afirmou que era pouco clara a fronteira entre os objetivos e obrigações do PENSA e os objetivos e obrigações de outro órgão da prefeitura, o Instituto Pereira Passos (IPP) – o que não deveria ocorrer se o objetivo estabelecido fosse claro e bem comunicado.

É bom você olhar lá porque ali está o objetivo expresso mesmo e eu acho que o principal é isso, usar os dados, as informações geradas pela prefeitura, para propor políticas públicas (Funcionário E5 do PENSA sobre o estabelecimento de objetivos).

Sim. Assim no nível top lá não teve conflito não, mas embaixo às vezes ficava um pouco qual era a fronteira entre o que o IPP faz e o que o Pensa faz (...) essa intercessão às vezes é meio complicada na definição dele que é até aonde vai a responsabilidade de um e de outro (Funcionário E5 do PENSA quando questionado se houve conflito com outros departamentos ou secretarias)

De acordo com a literatura acadêmica, outros motivos pelos quais o alinhamento estratégico é candidato a fator relevante na implementação do *Big Data* são, primeiro, o fato de ele possibilitar à organização identificar quais necessidades serão atendidas (DOMINGUES, 2013; MINELLI, CHAMBERS e DHIRAJ, 2013) e, segundo, o fato de permitir o desenvolvimento de capacitações para lidar com possíveis desafios futuros,



tendo em vista que os projetos de *Big Data* podem ser complicados, idiossincráticos e disruptivos (YAN, 2013; KAISLER, ARMOUR, *et al.*, 2013).

De fato, as entrevistas corroboram com a teoria nesses pontos, conforme pode ser observado nos trechos destacados a seguir. Por um lado, os funcionários das áreas clientes afirmaram ser a análise de dados uma necessidade para a execução de um bom planejamento e trabalho. Por outro, um dos funcionários do PENSA afirmou ser o departamento capaz de responder à mudança de preocupação do cidadão.

Nós [...] nunca tivemos a oportunidade de ter os dados e ainda mais de uma equipe especializada para nos ajudar a tratar desses dados. Então isso foi muito assim por ocasião de necessidade, precisamos fazer um estudo disso, a gente conversava com o Pensa [...] Nós trabalhamos com pesquisas, a secretaria reúne dados a partir de pesquisas e monta seus estudos e a gente consegue enxergar se uma linha está sobrecarregada ou não e aí fazer alguma ação pontual em cima daquela linha (Funcionário AC1 de uma das áreas clientes)

A gente precisa de uma informação que hoje a gente não tem. A gente precisa que essa informação seja disponibilizada de uma maneira mais clara. [...] Isso que a gente precisa, a gente precisa de pessoas que possam analisar informação em tempo real, então a gente está usando o PENSA para isso (Funcionário AC2 de uma das áreas clientes).

Não, não, eu acredito que tenha sido conduzido em função de algumas políticas. Então vamos supor, por exemplo, hoje qual é a preocupação do cidadão, a maior preocupação hoje? Mobilidade, mobilidade. Então nós temos que ter especialistas em mobilidade. Daqui a pouco vai mudar, não vai ser mais mobilidade, por exemplo (Funcionário E2 do PENSA).

Desta maneira, a análise do conteúdo das entrevistas apresenta evidências que elevam o alinhamento estratégico à condição de fator relevante na adoção e uso do *Big Data*, sobretudo no setor público.

### **6.1.2. Patrocínio da Alta Gestão**

Praticamente todos os atores envolvidos com a implementação do *Big Data* na prefeitura concordam com a importância do envolvimento da alta gestão, desde os funcionários da própria equipe do PENSA, até o representante da Casa Civil e os representantes das áreas clientes do PENSA. De fato, ao longo das entrevistas, o envolvimento da alta gestão pode ser percebido desde o início do planejamento, quando do estabelecimento de metas, até as atividades junto às secretarias, conforme pode ser visto pelos trechos destacados a seguir.

Eu diria que o prefeito dá a diretriz, o prefeito junto com a Casa Civil determina a diretriz e a equipe do PENSA, com base nessa diretriz, estabelece as suas prioridades. Então se você ver o acordo de metas da Casa Civil, você tem metas do PENSA dentro desse acordo. Então, o PENSA tem metas

específicas de entrega dentro do plano de metas da casa civil (Funcionário da Casa Civil).

Na realidade muitas das demandas que foram feitas inicialmente, foram feitas pelo prefeito, o próprio prefeito quis ter informações sobre tempo de espera em determinados pontos da cidade, tempos de viagem, que a própria Secretaria não tinha condições de gerar com aqueles dados de GPS. E aí o Pensa começou um estudo dentro dessa linha para o prefeito e se estendeu [...] nos direcionou bastante no planejamento do transporte desse evento (Funcionário AC1 de uma das áreas clientes).

A importância do envolvimento da alta-gestão na implementação do PENSA não se limita tão somente ao início das atividades. Conforme um dos funcionários do PENSA afirmou em uma de suas entrevistas, o processo de internalização de ideias depende de uma série de questões burocráticas e, por isso, o envolvimento de secretários e subsecretários é importante durante todo o processo.

Você tem um processo de internalização dessas ideias dentro da prefeitura que ainda depende de uma série de questões burocráticas que são inerentes a qualquer administração, por exemplo, não adianta nada a gente exibir quais são os locais que a gente mais problema com asfaltamento na rua se o orçamento daquele ano com os projetos para asfaltar já estão fechados, você vai ter que esperar para o ano seguinte [...] É fundamental – essa percepção – a gente vê quando a gente trabalha junto com os secretários (Funcionário E1 do PENSA).

Os trechos destacados corroboram com a literatura ao evidenciar a importância do *top management* no suporte e patrocínio da tecnologia durante sua implementação e fazendo a promoção do *Big Data* para garantir o seu uso. Yeoh e Koronius (2010) e Novo e Neves (2013) ainda destacam que o envolvimento da alta gestão também é importante para assegurar a disponibilidade e o acesso a todos os recursos operacionais necessários. De fato, pôde ser verificado que o envolvimento da alta gestão também foi importante nesse sentido nas entrevistas dos funcionários do PENSA, bem como na entrevista de um dos funcionários das áreas clientes.

Então ou eu dou sorte, não sei se é sorte, mas a gente começou muito alinhado porque como eu falei a ideia partiu do próprio prefeito e do secretário da Casa Civil. [...] A Casa Civil mexe com diversos órgãos e o prefeito é o prefeito então já começou de cima e isso facilitou bastante a integração e abertura de portas com os outros secretários e com as outras secretarias também porque tinha esse endosso (Funcionário E5 do PENSA).

Acho que é uma característica do Paes, de fazer uma gestão voltada para dados. E botar só secretários que também estão pensando sobre isso. Normalmente, costumamos acessar a alta gestão. Já chegamos na secretaria para falar com o secretário ou com o subsecretário, normalmente eles estão entusiasmados com a ideias e fazem acontecer. (Funcionário E2 do PENSA).

Entendo que se não vier pelo prefeito, pelo chefe da Casa Civil, as coisas não saem. Tem que ter o aval, a legitimidade deles, senão as coisas não saem da prefeitura. (Funcionário de uma das áreas clientes).

Na literatura acadêmica também consta que o envolvimento da alta gestão também é importante tanto na comunicação da estratégia de adoção da tecnologia, como na superação de barreiras à mudança e na definição das responsabilidades de cada envolvido com a implementação. Isso também pôde ser verificado nas entrevistas, tendo em vista que a alta-gestão solicitava estudos que só podiam ser realizados pelo PENSA e divulgava seus resultados às demais secretarias posteriormente. Isso pôde ser observado nas entrevistas do representante da Casa Civil e de um dos funcionários de uma das áreas clientes.

Eu acho que a gente na medida do possível, tenta arbitrar esse tipo de situação, e aí a alta administração onde coube, arbitrou e usou a informação que ajudou de uma forma mais correta a tomar a decisão, quem não disponibilizou dados, algumas situações [...] ajudava a liberar a informação (representante da Casa Civil sobre a atuação da alta gestão na solução de conflitos).

Não é que houve uma desconfiança do trabalho deles, é falta de conhecimento mesmo. Então assim, a primeira vez que a gente viu o relatório deles, foi uma questão do réveillon, que foi ótimo, ver todo mundo que está na praia de Copacabana meia noite, veio de Caxias, veio de São Gonçalo, superinteressante ver esse trabalho. Foi a 1ª vez que eu tinha visto. Então assim, não é nem uma desconfiança, que eles estão fazendo um negócio que não nos serve. A gente não sabe nem que potencial aquilo tem. Então assim, como a gente não sabe que potencial aquilo tem [...] Então não é uma falta de informação (Funcionário AC2 de uma das áreas clientes).

Desta maneira, a análise do conteúdo das entrevistas apresenta evidências de que o envolvimento da alta gestão é um fator relevante para a adoção do *Big Data*, sobretudo no setor público.

### **6.1.3. Alinhamento aos processos organizacionais**

Outro fator referenciado na literatura acadêmica como candidato à fator relevante para a adoção e implementação foi o alinhamento entre os processos organizacionais e de negócios e o *Big Data*. De acordo com o levantamento teórico, a compatibilidade da tecnologia com os processos organizacionais é um dos desafios da implementação.

De fato, se existe um departamento ou órgão responsável por alguma atividade específica ou cujo objetivo institucional está relacionado àquela atividade específica, então o uso de *Big Data* para estudos e projetos relacionados àquela atividade devem envolver o referido órgão ou departamento. Do contrário, pode surgir uma resistência à tecnologia. Tal qual como pode ser visto no tópico sobre o MODA, quanto menor o impacto nas tarefas cotidianas da organização, melhor. Essa questão foi levantada por um dos funcionários das áreas clientes, conforme trecho destacado a seguir.

Existe uma característica forte do prefeito de ser um cara muito centralizador e um cara que coloca a mão em tudo e ele sabe mesmo. Você pode perguntar para ele de qualquer projeto de qualquer secretaria, qualquer ação de qualquer secretaria, seja de saúde, educação, transporte, ele sabe, ele tem na palma da mão tudo que está acontecendo na cidade, todos os projetos que estão acontecendo. Então às vezes ele demanda ao PENSA diretamente independente da Secretaria [...] Não é conflito, a gente acha que o prefeito poderia demandar isso pela secretaria (funcionário AC2 de uma das áreas clientes).

Conforme já foi apontado por Markus (1983), no entanto, a resistência e o conflito não necessariamente são aspectos negativos da implementação de um SI. Podem representar resistências a alteração do *status quo* ou mudanças na estrutura de poder. Por vezes, desafiar a estrutura pode ser uma forma de se conseguir pensar diferente. E esse parece ser justamente o caso do PENSA, todos os entrevistados afirmaram não haver nenhuma alteração na relação entre secretarias e departamentos ou nos processos cotidianos com a criação do departamento de análise de dados, conforme trechos destacados a seguir.

Pela experiência que eu tenho, não houve grande mudança não. Não percebi isso não (Funcionário AC3 de uma das áreas clientes).

Fala de *Big Data*, a gente fala de implementação do plano estratégico, falou de *Big Data*, a interferência organizacional foi muito limitada, foi criada uma unidade a parte, quase um Pensa da prefeitura, que trabalha de forma axial com todo o resto do governo. Então uma interferência prática na estrutura, nenhuma, zero (Representante da Casa Civil).

Alguns entrevistados comentaram de uma alteração a nível mais operacional – mas não alterações que pudessem causar resistência à tecnologia. As alterações não implicaram novas atividades, apenas uma nova lista de prioridades ou reformulações de tarefas que não impactaram o *status quo* da secretaria, mas contribuíram para que houvesse maior aderência do processo às necessidades da população, conforme pode ser visto trechos destacados a seguir.

Eu acho que em alguns órgãos operacionais, mas assim não de estrutura você tem lá organograma dos órgãos... (Funcionário E5 do PENSA).

Não digo no processo não. Hoje a gente não tem um processo diferente do que a gente tinha, até por que é o que eu estou te falando, a gente pode encomendar estudos, mas eu não posso garantir que eu vou ter continuidade nesses estudos (Funcionário AC2 de uma das áreas de clientes).

Para as demandas do 1746 ela (a guarda municipal) funcionava muito sob demanda, quer dizer, tem um chamado de algum problema aqui, vou lá, resolvo e volto. Tem outro aqui, vou lá e volto. Tem outro lá, resolvo e volto. Esse estudo demandou uma mudança de operação na guarda que é fazer os guardas trabalharem, por exemplo, com roteiros, ao invés de ficar indo e atendendo ponto a ponto, eu vou ficar rodando aqui nessa área, que é a área que o estudo do PENSA me diz que é a área mais crítica da cidade e antes que o problema aconteça e eu tenha que ir lá [...] isso já diminui a possibilidade dele cometer a infração (Funcionário AC3 de uma das áreas clientes).

Além disso, um dos funcionários destacou que o PENSEA foi criado com o intuito de comportar uma abordagem mais aberta à exploração e experimentação da análise de dados, confirmando que no caso em questão o conflito destacado não era um aspecto negativo. Desde o princípio houve interesse que o PENSEA fosse um mecanismo que evitasse o engessamento em relação à descoberta e análise de dados, conforme trecho a seguir.

A primeira conversa que tivemos com o prefeito... A gente perguntou: O que você espera da gente? O que você espera que a gente comece a trabalhar primeiro? Ele falou: Não vou falar nada. Eu quero que vocês explorem, quero que vocês juntem tudo aí e vejam o que sai da cabeça de vocês, o que sai dos dados, façam análises e conforme forem aparecendo as demandas das cidades, dos cidadãos, vocês vão gerando produtos mais interessantes. Então assim tinha muita liberdade (Funcionário E5 do PENSEA).

A integração dos silos de dados é outro aspecto que torna o alinhamento com os processos de negócios um fator relevante na adoção e uso do *Big Data*. Para os dados serem úteis eles precisam atravessar as barreiras organizacionais, o que é, geralmente, difícil e causa brigas e disputas internas por poder corporativo (NOVO e NEVES, 2013). De fato, a partilha de dados dentro de um país entre os diferentes departamentos e agências do governo é vista como um desafio (KIM, TRIMI e CHUNG, 2014).

Essa preocupação com o silo do dado teve ressonância nas entrevistas. Conforme pode ser observado nos trechos a seguir, destacados das entrevistas do representante da Casa Civil e de dois dos funcionários do PENSEA.

Acho que não só em governo, acho que na vida, na iniciativa privada a mesma coisa, informação é poder, quem detém a informação tem poder. Então no momento em que você tem um órgão de cientistas, uma estrutura de cientistas de dados, que passam a solicitar os dados que você tem para transformar esses dados em informação, você no fundo está tirando o poder de determinadas estruturas da sua organização (Representante da Casa Civil quando questionado sobre silos de informação).

Tem algumas pessoas que gostam de reter informação, sempre tem esse perfil de gente (Funcionário E4 do PENSEA sobre silos de informação).

Qual o grande problema que o PENSEA enfrenta hoje? Primeiro é conseguir tirar os dados das secretarias. Porque como esse setor é novo a gente chega para você na secretaria de saúde diz “olha, nós queremos os dados da secretaria de saúde para cruzar com os dados da secretaria do meio ambiente para verificar se aquela quantidade de caminhões que a OCE está dizendo que passa...”, já é uma terceira fonte de informação. Então eu estou unindo 3 fontes distintas que justamente o setor foi criado para isso. Porque o meio ambiente não iria nunca se comunicar com a saúde que ia se comunicar com o consorcio das câmeras da OCE (Funcionário E2 do PENSEA).

No entanto, não houve problema específico para acessar um dado, como pode ser observado pelos trechos de entrevista a seguir.

Em geral os órgãos são bastante colaborativos nesse aspecto, a gente vê, quando a gente começa a falar com os técnicos, às vezes, é claro, que o cara tem aquela desconfiança: “espera, para que é isso?”. Aí você mostra os trabalhos que estão sendo feitos, apresenta para ele as propostas, em geral, dali para frente vai supertranquilo porque é algo que agrega para todo mundo (Funcionário E1 do PENSA).

Não existe isso, inclusive hoje dentro da arquitetura de sistema que nós montamos para os consórcios todos os dados ficam no IPLAN e o PENSA tem acesso a todos os dados, todas as informações porque nós deixamos. Então hoje não existe o dono do dado (Funcionário da AC1 de uma das áreas clientes).

A falta de conflitos relacionados à propriedade do dado não significa que a integração dos silos de dados não seja um fator relevante para a adoção do *Big Data*. O caso do PENSA, revela, na verdade, exatamente o oposto. A preocupação inicial com a integração dos silos de dados foi tão grande que deu origem ao decreto nº 37.541, de 13 de agosto de 2013, que permite que o PENSA tenha acesso a qualquer dado – o que talvez justifique a ausência de conflitos nesse tema –, conforme pode ser observado pelos trechos destacados das entrevistas de funcionários do PENSA.

Acho que o PENSA tem um decreto de liberação de dado interno. A princípio sim (acesso a todo e qualquer tipo de dado). Como normalmente não faço essa parte de interação de dados internos, trabalho mais com a parte de dados externos abertos, eu não encaro muito isso. Mas acontece, por exemplo, de ter dado de muito má qualidade e não temos como saber se o dado vem de lá de dentro mesmo, não sei, de alguma forma a secretaria não manda o melhor dado que ela tem (Funcionário E3 do PENSA).

Assim em relação aos dados não como eu já disse antes a gente não teve problemas. A gente tem esse decreto aí que foi criado que dá superpoderes de acesso a todos os lugares, mas a gente nunca precisou usar. Eu sei que existe isso, sei que outros órgãos... não sei é porque a gente está próximo do prefeito ou secretário e isso abre mais as portas... Não sei se tiveram facilidade... mas a gente não tem não (Funcionário E5 do PENSA sobre silos de informação).

Por fim, vale destacar de acordo com a teoria, o *Big Data* pode impactar a organização com o surgimento de uma cultura arrojada, meritocrática e de compromisso, que fica evidente pela maior atratividade pelos dados e os *insights* oriundos de sua exploração, maior cuidado com os modelos de análise – que passam a ser considerados de propriedade intelectual –, e o surgimento de um maior conforto na tomada de decisão com base nos dados (SCHMARZO, 2013; DAVENPORT, 2014). O surgimento dessa cultura pode ter impactos relevantes nos processos de negócios, o que eleva a importância do alinhamento dos processos organizacionais com o *Big Data*.

Esse ponto, no entanto, não pode ser verificado porque ao mesmo tempo em que se observou nas entrevistas a questão da cultura de dados como emergente, se observou que o Rio de Janeiro possui uma cultura voltada para a informação há algum tempo. Por

suposto, é inviável afirmar que o surgimento de uma cultura *data-driven* seja consequência da implementação do *Big Data* e do PENSEA. O que se constatou foi a ampliação desta cultura a partir de algum grau já iniciado de maturidade em análise em dados, conforme Schmarzo (2013) indicou que aconteceria.

De uma certa maneira sempre existiu essa cultura de dados e informações aqui dentro [...] essa cultura de dados, eu cheguei aqui em 2009, as vezes a gente chega e acha que tem tudo novo, que a gente é muito bom, e, na verdade, tem muita coisa sendo construída a muito tempo e eu acho que essa cultura de dados e de informação da prefeitura, é uma cultura que tem início lá nesse trabalho, nessa fundação do IPP, o IPP cria a mais ou menos 12 anos atrás o portal de informações extremamente sofisticado onde você consegue ter informações sobretudo o que você puder imaginar de dados relevantes da prefeitura [...] E aí dando sequência a essa cultura de informação de dados, a gente teve em 2010 um marco importante para a criação do centro de operações da prefeitura para a melhoria do funcionamento da cidade. O Centro de Operações (COR) ele foi inaugurado no final de 2010, extremamente tecnológico, centraliza uma série de órgãos da prefeitura dentro daquele meio, [...] passa a se produzir uma cacetada de informação [...] E eu acho que o terceiro movimento importante, eu estou tentando colocar isso em uma ordem lógica, foi a criação do PENSEA [...] Além de ter toda essa cultura de informação de massa de dados que a gente tem desde lá da criação do IPP, a criação do COR virou uma oportunidade impar porque passa-se a produzir informações online da cidade com uma massa absurda de dados. Então tem uma serie de informações ali, o PENSEA foi criado no final de 2013, então você tinha as informações do COR, tinha aqueles dados do IPP, você tinha uma serie de informações espalhadas aí de diferentes departamentos... (Representante da Casa Civil sobre a cultura de dados).

Eu acho que principalmente nessas secretarias que eu citei antes foi sendo criado mais um espírito assim de fazer políticas públicas de uma forma que você consiga medir como você está implementando suas ações. E cada vez mais não fazer tantas coisas por intuição ou só baseado na experiência técnica que é importante, mas é importante também ter uma visão dos dados aí até para justificar o que está sendo feito (Funcionário E5 do PENSEA).

Tendo a pensar que isso vai acontecendo naturalmente. Só de começarmos a demandar dados, demandar análise, as secretarias [...] prefeitura começa a convergir para uma gestão mais voltada para os dados (Funcionário E1 do PENSEA).

As entrevistas apresentaram evidências de que o alinhamento dos processos organizacionais e de negócios é sim um fator relevante na implementação e uso do *Big Data*.

## 6.2. FATORES RELACIONADOS AOS RECURSOS HUMANOS

Outro fator candidato à relevante para a implementação e uso do *Big Data* é o envolvimento de funcionários qualificados. Conforme bem resumiram Minelli, Chambers e Dhiraj (2013), “a tomada de decisão baseada em dados é uma jornada, e

sem o talento certo ao longo dos níveis organizacionais, os benefícios da ciência da decisão não podem ser efetivamente realizados”. Esse fator é considerado um desafio, sobretudo devido à falta de profissionais com as habilidades e conhecimentos necessários para lidar com a análise dos dados típicos da era do *Big Data*.

De fato, são tantas as capacitações e competências que um cientista de dados deve ter (MINELLI, CHAMBERS e DHIRAJ, 2013; DAVENPORT, 2014) que não é difícil de entender porque há uma enorme dificuldade em encontrar alguém com todas as habilidades descritas (DAVENPORT e PATIL, 2012).

A dificuldade de encontrar um Cientista de Dados no mercado é ainda maior quando se trata de um emprego no setor público. Conforme um dos funcionários de uma das áreas clientes comentou em sua entrevista, o setor público demanda um tipo de profissional que valorize mais uma carreira de longo-prazo do que uma de curto-prazo.

Qual é a questão? Para você ter um profissional desses, é um profissional que está sempre se atualizando, são pessoas muito jovens, enfim, a galera que põe a mão na massa mesmo, fica o dia inteiro ali analisando dado. Então assim, como você mantém um profissional desses dentro da prefeitura? A prefeitura é uma carreira de longo prazo, depende do momento político, então tem momentos políticos que isso vai ser supervalorizado e em outros momentos não. Então a gente nunca sabe, política é isso, nunca sabe o que pode acontecer (Funcionário AC2 de uma das áreas clientes).

Conforme pôde ser visto na revisão bibliográfica, várias soluções têm sido propostas para resolver a crise de talentos (YAN, 2013). No entanto, ao considerar o perfil do profissional que trabalha no setor público, é fácil perceber que nem todas as soluções propostas são adequadas. De fato, as únicas soluções viáveis para resolver a crise de talentos no caso do PENSA são o investimento formal através de instituições de ensino superior e a formação interior de talentos.

O decreto de criação do PENSA, decreto nº 37.215, de junho de 2013, que especifica a criação de “convênios com universidades, instituições e empresas quando se fizer necessário e que tragam conhecimento e inovação para os processos em questão” revela que a prefeitura estava atenta a isso. É válido, no entanto, destacar que o desenvolvimento desses convênios para investir em uma educação formal por meio de instituições de ensino superior, como descrita, é uma solução voltada para o longo-prazo. Com efeito, para o início das atividades do PENSA, a prefeitura buscou encontrar funcionários com o perfil adequado no ensino superior, como indicou a entrevista dos funcionários do PENSA.



Eles (os primeiros funcionários) vieram com o folder que foi colocado dentro das universidades, inclusive na COPPE, solicitando que essas pessoas viessem para a entrevista (Funcionário E2 do PENSA sobre os primeiros candidatos externos para as vagas para funcionários).

A primeira seleção foi feita a partir de uma chamada com uma série com uns... Foram uns 200 currículos que a gente recebeu para analisar e a gente selecionou essa equipe inicial (Funcionário E1 do PENSA).

A formação interior de talentos, a outra solução viável de acordo com a literatura acadêmica, também teve destaque na concepção do PENSA. O decreto de criação do PENSA mostra que uma das competências do órgão, como instituição é “formar profissionais da área de *Big Data*”.

De fato, o PENSA fez uso de treinamentos para desenvolver as próprias competências, além de estimular um desenvolvimento de competências por parte dos seus próprios funcionários por meio da transferência de conhecimentos entre eles. A equipe original do PENSA, por exemplo, teve treinamento, como pode ser visto nos trechos destacados de entrevistas com os funcionários do PENSA abaixo.

Houve pagamento de tudo para todos aqui, curso da *Big Data* da PUC, da FGV, todos fizeram os treinamentos, mas o que a gente percebe é que de fato é que esse estudo é tão rápido, é tão novo que umas das características que a pessoa tem que ter é a capacidade de se atualizar sozinho pela internet fazendo curso online, pegando documentação que está saindo, participando de fóruns (Funcionário E1 do PENSA).

Cada um tomava sua iniciativa, a gente conversava e beleza. E de treinamento da equipe original todo mundo. Todo mundo fez um curso de *Big Data* mesmo na PUC ou na FGV. Mas também tinha muito curso que a gente dava um para o outro ali no dia a dia. Me ajuda aqui? Você sabe fazer isso? Então passava o conhecimento de um para o outro (Ex-funcionário do PENSA).

A troca de conhecimento entre os integrantes da equipe, no entanto, não ocorre apenas por treinamentos. Pelo contrário, há também processos informais de passagem de conhecimento bastante eficazes, como pôde ser observado nas entrevistas de funcionários do PENSA.

Uma equipe pequena assim, eu acho que as coisas não precisam ser tão formais, você tem que ter uma estrutura certamente. É, tudo é sobre demanda, você vai fazer alguma que não tem demanda, é um hobby, você está fazendo um hobby (Funcionário E4 do PENSA sobre processos de treinamento e aprendizado de novas competências).

Bom, até agora foi completamente *peer-to-peer*, não sei fazer o script para salvar dados do Twitter [...] Não teve nenhum treinamento formal. Mas parece que agora o pessoal [...] (vai) passar um treinamento para a gente de integração de dados. É porque normalmente, todo mundo na equipe tem uma característica de Hacker bem forte, então não temos muito problema de aprender on-line (Funcionário E3 do PENSA sobre os processos de treinamento e aprendizado de novas competências).

Sim. Hoje em dia todo mundo lá virou um processador, todo mundo sabe mexer um pouquinho. Eu melhorei muito na parte de estatística, na parte de banco de dados também. Então tem isso assim eles botam lá uma ferramenta e o pessoal gosta [...] Eu aprendi muito enquanto estava lá. Acho que consegui ensinar um pouco também porque senão ficaria completamente cada um setorizado e ficaria uma linha de montagem. Então tinha que ter essa mistura até para ter essas ideias inovadoras, precisava de um mix ali do conhecimento técnico e cada um e experiência anterior também até porque além de cada um ter uma formação fora da prefeitura teria que conseguir agregar coisas diferentes (Funcionário E5 sobre a troca de conhecimentos).

Tendo em vista as dificuldades relacionadas à contratação e à formação de um cientista de dados, a prefeitura, no momento da criação do PENSA, adotou como abordagem aquela destacada pela literatura acadêmica como alternativa óbvia: a montagem de uma equipe cujos funcionários tenham todas as habilidades necessárias em conjunto, sem que cada um dos funcionários tenha necessariamente que saber todas as habilidades (DEVLIN, ROGERS e MYERS, 2012; DHOLAKIA e DHOLAKIA, 2013; DAVENPORT, 2014).

O processo de formação da equipe pôde ser verificado em diversas entrevistas, dentre as quais destaco um trecho da entrevista do representante da Casa Civil.

No processo de construção do PENSA, foi uma equipe de 5 a 6 pessoas, um líder e mais 6 ou um líder e mais 5, e aí a estratégia foi trabalhar em duas frentes, de um lado, buscar dentro da máquina pessoas com esse perfil de cientistas de dados que a gente não sabia exatamente o que era esse perfil, mas quando a gente colocou na descrição do cargo para ocupar essa posição, tiveram pessoas com formação acadêmica extremamente sólida, facilidade em trabalhar com números e pessoas com uma capacidade analítica muito grande, pessoas capazes de solucionar um problema e ao mesmo tempo trazer soluções fora da caixa, esse foi uma frente de trabalho. E a outra frente foi buscar nas universidades, foi feito um trabalho nas universidades para buscar essas pessoas direto Lá da academia (Representante da Casa Civil sobre a formação da Equipe).

É importante destacar que a prefeitura não buscou um cientista de dados e funcionários complementares para alavancar seus esforços. Ao invés disso, buscou funcionários que eram qualificados em análise de dados e que se complementassem no intuito de desenvolver análises de dados com qualidade e que gerassem *insights* interessantes, conforme pode ser visto nos trechos destacados a seguir.

Então essas coisas de *Big Data* também são complicadas de ver onde começou ser chamado de *Big Data* e onde era sei lá análise de dados. Então todo mundo tinha um background de análise de dados, mas assim ser *Big Data* ou não eu não consigo afirmar. [...] No PENSA a gente teve seis pessoas inicialmente mais o gestor cada um de uma área o que eu acho que ajuda porque como não tem um perfil certinho de cientista de dados eu acho que foi boa essa mistura porque conseguiu juntar um pouco o que a equipe realmente precisava e não os indivíduos (Funcionário E5 do PENSA sobre a formação da equipe).

Primeiro, todo mundo que está aqui de alguma forma, sempre se interessou por dados, não tem ninguém formado em ciências de dados. Normalmente recorremos a isso, o físico que trabalha bastante com dados durante a faculdade, eu que fiz meu mestrado em análise de dados pesado. Mas pensar em tudo o que um cientista de dados precisaria para trabalhar é bastante coisa e acho que ninguém da equipe do PENSA tem tudo isso. Aí vamos fazendo exatamente como você falou: uma integração de equipe (Funcionário E3 do PENSA).

Segundo Yeoh & Koronius (2010), a equipe deve ser multifuncional e composta por ambos os técnicos e de negócios, tendo em vista que a equipe tem que lidar com diversas plataformas e interfaces, conexões com sistemas legados, diversas ferramentas, em suma atividades que exigem diferentes habilidades e competências. Um mix adequado de conhecimentos técnicos e de negócios é a chave para o sucesso. A prefeitura procurou considerar isso na formação da equipe, como pode ser observado pelas entrevistas do representante da Casa Civil e dos funcionários.

Aqui para dentro a gente conseguiu duas, um PhD em matemática e um mestrando em geografia, skills bem interessantes, e buscou mais três pessoas, tinha um PhD da COPPE que trabalhava com neurociência, o outro era de engenharia de computação, outro cara de engenharia de computação e um outro cara que tinha trabalhado com educação e gestão. Então era uma turminha, todos eles muito bons de número, extremamente analíticos e ao mesmo tempo com informações muito distintas o que, de uma certa maneira, nos agradava porque eram pessoas sem os vícios que a gente já tem da administração pública. Informações bem complementares (Representante da Casa Civil).

Ele decidiu pegar 2 pessoas da própria estrutura da prefeitura no caso, eu vim da educação e o X, que é geógrafo. A ideia era montar um grupo totalmente multidisciplinar cada um com conhecimento específico para poder haver uma convergência na hora de analisar os dados [...] Então hoje cada um tem seu nicho de que trabalha, por exemplo, eu agora estou num projeto de redução de consumo de água em escolas que é um de educação. Então eu tenho uma facilidade, eu conheço as pessoas da educação. Então tem uma pessoa que é especialista nessa parte (Funcionário E2 do PENSA).

Sim. Mas todos são da área de licenciatura. Granja e eu somos engenheiros, Sérgio físico, Léo cientista da computação e o Pablo que é advogado, mas é um hacker e está fazendo doutorado em Engenharia agora. Isso é uma coisa que sentimos falta, temos falado em trazer algum arquiteto para cá (Funcionário E2 do PENSA quando questionado se a equipe era multidisciplinar).

Cada um tem uma especialidade, eu acho que isso aí é uma coisa válida certamente, a divisão de tarefas é uma coisa que acontece naturalmente. É difícil achar uma pessoa que sabe tudo, mas achar 3 pessoas que sabem, uma coisa de cada um, e esses 3 conseguem conversar, é muito mais fácil (Funcionário E4 do PENSA)

Nesse sentido, a única recomendação da literatura que não pode ser verificada na formação da equipe do PENSA foi a constituição de uma equipe de apoio à equipe de *Big Data* com funcionários dotados de competências específicas que contribuam com o

aumento de produtividade, como, por exemplo, analistas de bancos de dados (DAVENPORT, 2014). Isso foi inclusive ressaltado por um dos entrevistados.

Cara eu acho que tem que ter alguém um gerente de banco de dados uma coisa que a gente nunca teve que atrapalhava muito e até hoje não tem, um DBA mesmo. (Funcionário E5 do PENSA sobre a possibilidade de um profissional de apoio e conhecimento técnico).

Perfil técnico que mais falta aqui é um DBA, alguém que realmente entende como estruturar os dados, uma coisa que a gente falava muito desde o começo, precisamos de um DBA [...] realmente um DBA seria bem-vindo, é a época que a gente tentou pegar um (cedido) pelo IPLAN [...] mas é porque parece que o IPLAN cortou (Funcionário E4 do PENSA sobre a possibilidade um profissional de apoio e conhecimento técnico).

É interessante destacar que tão importante quanto contratar os integrantes da equipe é a necessidade de identificar um gerente de projeto e especificar papéis e responsabilidades de cada membro da equipe. Os gerentes da equipe de análise de dados também desempenham um papel de liderança e de construção de cultura no que diz respeito à *Big Data* (DAVENPORT, 2014; OHLHORST, 2013). A importância de uma pessoa responsável por identificar as responsabilidades de cada membro da equipe também pode ser observada nas entrevistas.

Na gestão tem que ter alguém que seja bom tecnicamente como politicamente e também bom gerencialmente porque assim você coloca diversas informações então tem que ter alguém ali que bota o pessoal para trabalhar porque senão não funciona então as informações não vão fluir de cima para baixo nem de um lado para o outro. [...] A questão de programação também é importante. (Funcionário E5 do PENSA sobre a composição da equipe exigir um responsável que consiga se comunicar com os altos gestores e com os analistas de dados).

Vale ainda destacar que a comunidade acadêmica afirma que a equipe de *Big Data* não deve ser de natureza estática e deve ser capaz de evoluir e adaptar-se às necessidades do negócio, independente da estrutura adotada. A equipe deve sempre ter em seu *staff* as competências e capacitações necessárias para realizar um projeto antes de aceitá-lo. Essa preocupação também pode ser observada na entrevista de um dos funcionários da equipe, conforme pode ser observado no trecho destacado a seguir.

Todas essas tecnologias no começo a gente tinha um trabalho muito grande de identificar, mapear e trazer junto com a estruturação do servidor, junto com a estruturação da coleta de dado, isso eu diria que foi mais ou menos, junto a produção de ideia. A gente tinha que começar a fazer tudo ao mesmo tempo e o primeiro ano foi muito disso de estruturação e etc. A partir agora desse segundo ano é que a gente está com a capacidade de começar a coletar, pegar dados que a gente já tem no banco e começar uns estudos de formas mais rápidos. Então muda um pouco até o perfil da equipe, é o perfil que a gente precisa, e eu tenho certeza que daqui 1 ano as coisas vão ser diferentes (Funcionário E1 do PENSA acerca da composição da equipe).

Vale lembrar que, de acordo com a revisão bibliográfica, tendo em vista as dificuldades de contratar um profissional talentoso em análise de dados, as empresas devem esforçar-se de modo a reter os profissionais já contratados. Em outras palavras, reter bons profissionais é tão importante quanto contratar funcionários adequados. De fato, a prefeitura teve que lidar com esse desafio, à medida que os integrantes da equipe inicial foram deixando de ser funcionários da prefeitura e sendo substituídos.

Engraçado que daquela configuração inicial daquelas 6 pessoas hoje temos só 2 (Funcionário E2 do PENSA).

De você encontrar e você reter. A gente perdeu a maior parte da nossa equipe. Era uma equipe de 7 pessoas (Funcionário E4 do PENSA).

Ficou claro também que a transferência de conhecimento dentro da equipe atenuou o impacto da falta de sucesso na retenção de funcionários, conforme pôde ser destacado das entrevistas dos funcionários da equipe do PENSA.

(A equipe) diminuiu [...] Antigamente tinha X (que era um geógrafo), ele trabalhava bastante com mapa, hoje em dia sou eu quem faz isso. Quando alguém precisa mapear ou fazer uma análise espacial, eles falam comigo. (Funcionário E3 do PENSA).

Quando questionados sobre os principais motivos que levavam ao *turnover*, os funcionários levantaram diversos motivos, o que reforça a ideia de que a retenção é um desafio. Dentre os motivos levantados, pode-se citar a falta de oportunidades de crescimento, a visibilidade de mercado e a desmotivação, como pode ser observado nos trechos destacados a seguir.

Cara eu saí por diversos motivos. Um deles foi você ficar dois anos lá trabalhando como técnico, eu sou formado mais para a parte de gestão e lá eu não ia conseguir evoluir [...] (E existiram outros motivos?) Eu acho que mais de relacionamento interno de lá em geral assim porque eu estava me desgastando. Muitas pessoas saíram por isso também [...] Se ainda eu visse para onde eu iria evoluir lá tudo bem. O que me desmotivou mais foi que eu precisava crescer e lá não tinha como evoluir mais (Funcionário E5 do PENSA).

Engraçado que daquela formação inicial, daqueles seis, hoje só temos dois. Quatro pessoas já saíram, saiu e isso é uma outra coisa ruim nesse mercado, porque as pessoas começam e ganham visibilidade, começa uma coisa nova, outras instituições que estão criando e não querem começar do zero e acabam entrando em contato com essas pessoas e assim o grupo vai se desfazendo e foi conseguindo se refazer (Funcionário E2 do PENSA)

Dificuldade pode acontecer que, quanto mais gente você coloca em uma equipe, mais dificuldade você tem de coordenação, mas quando a equipe é pequena, uma equipe de 7 pessoas não tem dificuldade de comunicação, no máximo pode ter picuinha, mas dificuldade de comunicação não tem [...] A equipe ficou desmotivada, mas isso é outra questão, realmente foi problema de gestão (Funcionário E4 do PENSA).

Com relação às abordagens exatamente sugeridas por Davenport (2014), vale destacar que não foi possível, no entanto, verificar se a recompensa financeira pode contribuir para a retenção de um funcionário na equipe, tendo em vista que esse tipo de recompensa tem sido utilizado pela prefeitura desde antes da adoção do *Big Data*, conforme pode ser visto no trecho destacado a seguir.

Isso tem para todo mundo que é da prefeitura. A prefeitura tem as metas de cada órgão, tem as suas metas gerais do planejamento estratégico e todo ano essas metas são divididas em todos os órgãos, todas as secretarias, todos os órgãos têm as suas metas. Então a prefeitura inteira tem metas e quem bate as metas ganha um variável [...] (Qual foi a importância disso para o Pensa?) Cara... Quem entrou no ano que foi criada a equipe já entrou no período depois do que poderia se candidatar para ganhar do ano anterior. Então só agora que vai então eu não tenho muito que falar disso não, não deu tempo. É porque a gente saiu também então não deu tempo de medir (Funcionário E5 do PENSA).

Por outro lado, foi observado que um dos fatores que foi levantado como motivo para o *turnover* – o reconhecimento –, pode ser visto também como uma recompensa não financeira, dependendo da maneira como for trabalhado pela alta-gestão, conforme foi destacado por um dos entrevistados.

Internamente eu acho que isso nunca foi muito bem estruturado não como deveria ser. Mas uma coisa que poderia melhorar lá, uma coisa que ajuda muito, que recompensa muito às pessoas é a visibilidade que tem uma penetração muito grande no auto escalão que dá visibilidade para as pessoas que trabalham lá, você pode estar apresentando para quem vai te contratar daqui a pouco a promoção (Funcionário E5 do PENSA).

De maneira análoga, é importante destacar que não pode ser verificada como fator relevante a garantia de trabalho com outros cientistas de dados, de modo a estimular intelectualmente e aumentar as chances de retenção deles, conforme sugerido por Davenport (2014), pois o PENSA optou pela abordagem de equipe e não por contratar cientistas de dados (DAVENPORT, 2014).

Por fim, vale destacar que embora não tenha sido observado na literatura, um resultado interessante das entrevistas foi verificar que a gestão de expectativas dentro da equipe de analistas de dados é um fator importante na motivação dos funcionários e, por conseguinte, pode contribuir para sua respectiva retenção.

Então eu acho assim, se você está em um setor privado, e você vai fazer um negócio, como o dinheiro é um pouco mais contado, as pessoas sabem o que é que eles estão esperando, não que aqui não tenha tido uma expectativa, mas a expectativa daqui foi publicitária, sabe, se acontecesse alguma coisa, além disso, era efeito colateral, mas não era realmente prioridade, isso desmotiva muita gente da equipe daqui, e além disso muito termo de entrevista não foi cumprido, sabe, não deveria ter precisado ir para o escritório. De fato eu tenho um horário muito flexível, mas eu não precisaria ter ido para o

escritório, iria ser focado em resultado, não iria ser focado em meios, isso não aconteceu, e por causa disso a equipe toda saiu sabe, quem sobrou ficou por questões externas, eu estou aqui por questões externas (Funcionário E4 do PENSA).

Isto posto, o presente trabalho pode verificar que há evidências que o envolvimento de funcionários qualificados é um fator relevante na adoção e uso do *Big Data*. Nesse sentido, foi constatado nas entrevistas que a abordagem de equipe é uma solução tendo em vista a escassez de funcionários dotados de todas as competências necessárias a um cientista de dados. Além disso, foi possível perceber que tão importante quanto à contratação de bons funcionários para a formação da equipe, é a retenção dos funcionários. A retenção foi um dos grandes desafios enfrentados na formação do PENSA. Assim, com base nos resultados descritos até aqui, considerando também que este trabalho é uma pesquisa exploratória, faço as seguintes proposições.

<b>Proposição 1</b>	<b>Quatro fatores organizacionais são relevantes na implementação e uso de <i>Big Data</i> no setor público: o alinhamento à estratégia da organização; o patrocínio da alta gestão; o alinhamento entre os processos organizacionais e o sistema de <i>Big Data</i>; e o envolvimento e retenção de funcionários qualificados em análises de dados.</b>
1.1	O alinhamento estratégico da implementação do <i>Big Data</i> é relevante na adoção e uso do sistema no setor público, pois identifica quais necessidades serão atendidas, quais capacitações devem ser desenvolvidas e estabelece objetivos claros, mensuráveis, realizáveis e verificáveis, norteando e sustentando o compromisso de todos.
1.2	O envolvimento da alta gestão na implementação e uso do <i>Big Data</i> é importante tanto para fins de suporte e patrocínio à nova tecnologia, como para fins de comunicação da estratégia de adoção e auxílio na superação de barreiras à mudança.
1.3	O alinhamento entre os processos organizacionais e o <i>Big Data</i> é um fator relevante na implementação e uso no setor público, pois diminui o impacto da tecnologia no <i>status quo</i> e a resistência à implementação, além de auxiliar no desenvolvimento de uma cultura mais aberta à exploração dos dados e na superação dos desafios oriundos da integração dos silos de dados.
1.4	O envolvimento e retenção de funcionários qualificados em análises de dados é um fator relevante na adoção e uso do <i>Big Data</i> no setor público. A busca de profissionais nas instituições de ensino superior e a formação interior de talentos são soluções para o <i>Talent Gap</i> no médio e longo prazos. No curto prazo, a abordagem de equipe é uma solução eficaz para reunir as competências necessárias para execução de análises de dados.

## 6.3. FATORES RELACIONADOS À IMPLEMENTAÇÃO

### 6.3.1. Estruturas facilitadoras da Implementação

De acordo com a academia, a utilização de estruturas e funções específicas para a implementação – tais como comitê gestor, líder de projeto, campeão de projeto e grupos de superusuários – podem aumentar a probabilidade de sucesso da adoção do *Big Data*.

O comitê gestor garante que o desenvolvimento da implementação ocorra como planejado, tendo em vista que sua aprovação faz-se necessária em cada etapa da implementação – o que garante que existam controles adequados sobre os processos de tomada de decisão da equipe e assegura de maneira eficaz a participação adequada dos principais agentes relacionados ao sucesso do sistema –, bem como permite que a alta administração monitore o progresso do projeto (SOMERS, NELSON e RAGOWSKY, 2001).

Nada foi comentado a respeito de comitê gestor em nenhuma das entrevistas, o que, em um primeiro momento, poderia levar à crença de que essa estrutura não é relevante para o processo de implementação e uso do *Big Data*. No entanto, conforme foi discutido anteriormente, a prefeitura teve uma postura proativa e estabeleceu em decreto que o PENSA teria acesso à todos os dados da prefeitura. Além disso, o gestor foi escolhido justamente por ter facilidade de comunicação com a alta gestão e por ter conhecimento técnico suficiente para lidar com os funcionários e analistas de dados da equipe, conforme foi comentado em uma entrevista. Assim sendo, a implementação de um comitê gestor propriamente dito, considerando que a equipe do PENSA é relativamente pequena frente ao tamanho da prefeitura, seria mera formalidade. Por conseguinte, não é possível concluir nada a respeito da relevância do comitê gestor para a implementação e uso de *Big Data* a partir do caso estudado.

A literatura acadêmica também afirma que a existência de um líder/campeão de projetos pode aumentar as probabilidades de sucesso da implementação e do uso do *Big Data*, sendo, portanto, um candidato à fator relevante para a iniciativa de implementação. Entre as funções do líder/campeão de projetos estão a manutenção do contato com todos os usuários e monitoramento do uso e dos problemas do sistema. O líder deve ainda trabalhar no sentido de aumentar o comprometimento dos usuários e funcionários com o sistema de *Big Data*.



Como figura hierárquica, o líder/campeão de projeto deve ser um defensor do sistema e de seus benefícios, ajudar a resolver conflitos, gerenciar resistência e, entre outras competências, ser capaz de prever os desafios organizacionais e a mudança de curso no planejamento, conforme pode ser visto na revisão bibliográfica. Este líder não é a pessoa dotada de todos os conhecimentos técnicos e sim uma pessoa capaz de resolver as os problemas de implementação (LEININGER, 2015).

Pelas entrevistas, foi possível perceber que o PENSA possuía um líder de projeto, o seu gestor, conforme destacado na entrevista por um dos funcionários da equipe e por um dos funcionários de uma das áreas clientes.

Então, uma das coisas que fizeram a seleção do nosso gestor, era uma pessoa que sabia se apresentar bem na prefeitura. Teve muito projeto que nasceu na cabeça dele e ele tentou vender, nenhum aconteceu, porque é aquela coisa, a secretaria tem os problemas dela, se você chegar de fora impondo as coisas, até porque a gente não conhece os detalhes de como é que as coisas funcionam lá dentro, não acontece. É claro que a gente não vai fazer tudo o que eles pedem, a gente tem que conhecer o que está acontecendo lá para pensar: “Esse aqui pode ser um problema significante”. Eu prefiro saber o que está acontecendo e filtrar o que a gente pode contribuir, do que sair tentando vender uma coisa para ele não vai querer, porque não é a gente que vai fazer nada, são eles, não tem esse negócio de bater na mesa (Funcionário E2 do PENSA sobre a existência de um líder/campeão de projeto).

Teve, me parece que foi o Gestor (Funcionário AC3 de uma das áreas clientes quando questionado sobre o líder de projeto).

É uma pessoa que não só interage a equipe, mas também faz a gestão das expectativas. É uma pessoa que atua junto as outras áreas para garantir que nós não somos x9, nós só estamos aqui para ajudar, vocês tem muito a ganhar com a gente e a gente tem muito a ganhar com vocês (Representante da Casa Civil).

Na prática, alguns usuários das áreas clientes se tornam campeões de projeto, e começam a defender a iniciativa, como pôde ser observado na entrevista de um dos funcionários do PENSA.

Quando estamos conversando com a secretaria, alguém de lá se apaixona pela ideia e talvez vire esse campeão lá dentro. Mas isso nunca foi delimitado. Como dependemos muito da interação com as secretarias para implementar o projeto, a pessoa faz isso, ela passa confiança (Funcionário E3 do PENSA sobre líder/campeão de projeto).

Por conseguinte, as respostas me levam a crer que tanto um líder como um campeão de projetos são fatores relevantes na adoção e uso do *Big Data*, sejam ambos instituídos formalmente ou não, sejam funções desempenhadas por pessoas diferentes ou não.

Por fim, o grupo de superusuários aumenta a comunicação e a troca de conhecimento entre os usuários e melhora os resultados do treinamento, aumentando as chances de o

sistema ser utilizado como planejado. O principal benefício é o retorno às funções normais ao final da implementação, trabalhando ao lado dos usuários finais e garantindo assim uma maior eficácia do treinamento (SKOK e LEGGE, 2001).

Nas entrevistas, ficou claro que não houve um processo de escolha de superusuários.

Efetivamente não temos isso (Funcionário AC3 de uma das áreas clientes sobre Superusuários)

Alguns indivíduos desempenharam às vezes de superusuários, tanto nas áreas clientes como no próprio PENSA, levando em consideração que a necessidade de ambos era justamente a existência de um perfil capaz de comunicar-se com ambos em suas respectivas linguagens técnicas, e assim contribuir para o desenvolvimento de estudos melhores e mais rápidos, além de ser capaz de atender as demandas específicas com mais rapidez – a disponibilidade de um dado ou uma explicação específica da área para o PENSA ou a explicação de algo técnico para a área cliente. Isso pôde ser observado tanto em entrevistas com funcionários das áreas clientes na entrevista com um dos funcionários do PENSA.

O que existe é o seguinte, se concentrou mais em um técnico do Pensa nos projetos de setor de transporte e com certeza esse cara já tem mais conhecimentos da área de transporte, inclusive é a pessoa que está sendo alocada para a gente. Então ele hoje é a pessoa mais capacitada à entender e trazer um resultado mais rápido. Qualquer outro ia fazer um negócio “não bem assim” demoraria um pouquinho mais o desenvolvimento e ele talvez tivesse um pouco mais de capacidade de velocidade, não é nem capacidade. Capacidade de fazer uma coisa mais rápida (Funcionário AC1 de uma das áreas clientes sobre Superusuários).

Esse geógrafo seria como esse superusuários, de uma maneira espontânea ele serviu para nos mostrar o potencial. Porque ele tinha uma relação pessoal com um dos integrantes e sabia o que estava acontecendo lá (Funcionário AC2 de uma das áreas clientes sobre Superusuários).

Existe. É o cara que vai fazer tudo na verdade, só quando precisar de inputs externos ou talvez em algum momento desenhar a metodologia, ele senta com as outras pessoas e conversa. Mas quem toca o projeto mesmo é uma pessoa da equipe (Funcionário E3 do PENSA quando questionado sobre Superusuários).

Por conseguinte, as entrevistas levam a crer que a existência de superusuários pode sim alavancar as probabilidades de sucessos da implementação e uso do *Big Data*, sendo assim fator relevante para a iniciativa.

Portanto, o presente estudo encontrou evidências que lhe permitem afirmar que o uso de estruturas facilitadoras à implementação – comitê gestor, líder/campeão de projetos e grupos de superusuários – é relevante na adoção e uso de *Big Data*.

### 6.3.2. Programa de Implementação

Relata-se que os projetos de *Big Data* falham 30% mais frequentemente do que outros projetos de TI, e as principais razões relatadas para tal são relacionadas à gestão da implementação do *Big Data* (INSTITUTE FOR ADVANCED ANALYTICS, 2014). Desse modo, é de se esperar que o método de gestão do projeto de implementação seja um fator relevante a ser considerado. De maneira geral, um projeto de tecnologia da informação pode ser desenvolvido e implementado por diversos métodos: implementação em cascatas, progressiva, ágil, entre outros.

É, portanto, curioso que não tenha sido verificado nas entrevistas o uso formal de um método de implementação de projetos ou realização de estudos junto às secretarias, conforme pode ser observado nas entrevistas de todos os funcionários do PENSA, quando estes foram questionados sobre o estilo de planejamento.

Nós aqui no PENSA não utilizamos nenhuma metodologia de desenvolvimento de projetos, não temos nenhuma estrutura de projeto, voltado para projeto. Nós chamamos de projeto como todo mundo chama de projeto: ah, vamos fazer um projetinho de consumo de energia elétrica aqui na escola? O professor vai, escreve um roteirinho com os alunos, eles vão lá para medir quanto de energia elétrica aquele aparelho consome, quanto aquele dali, foi feito um projeto, mas nenhuma metodologia para projeto, nós não seguimos nenhuma metodologia específica (Funcionário E2 do PENSA).

Nada estruturado ainda, é uma coisa que vai acontecendo normalmente também (Funcionário E3 do PENSA).

No entanto, é importante destacar que essa ausência de um método de planejamento não foi vista como um desafio, conforme pode ser visto pela entrevista de um dos funcionários do PENSA, que afirmou ser de grande importância a ausência de um método de planejamento de projetos, pois dava maior liberdade para o desenvolvimento e finalização do projeto.

(Qual é a importância dessa ausência?) Imensa. Para organizar as coisas, os projetos tinham muita iniciativa e pouca “acabativa”. Então se você tem uma organização melhor você consegue entregar mais (Funcionário E5 do PENSA).

É interessante, entretanto, destacar que essa ausência de um método de implementação de projetos era na verdade algo bem próximo daquilo que a literatura classifica como método ágil de gestão de projetos. Mesmo os benefícios destacados pelos funcionários da equipe são justamente aqueles que caracterizam o método ágil – foco nos resultados, periodicidade de encontros alta, apresentação de resultados em ciclos rápidos e tempo de execução dos projetos pequeno –, como pode ser observado na entrevista de um dos funcionários de uma das áreas clientes e de um dos funcionários do PENSA.

(Existe algum cronograma?) 3 meses, eu acho. (Daqui a 3 meses eles entregam algum resultado?) É, já estão apresentando alguns resultados. Eles já estão passando, faz a primeira prévia, manda para a gente validar, a gente retorna com nossos comentários. Então é um processo... (Funcionário da AC2 de uma das áreas clientes).

Uma regra que a gente tem aqui passada pela Casa Civil: ela quer que a gente não seja burocrático, ela quer que a gente seja o mais solto possível. Então eu acho que isso é positivo por um lado, negativo por outro. Tem ganhos e tem perdas em cada um desses casos. No caso dessa gestão, a maior parte das demandas quando elas vem de fora, são coisas urgentes. Isso a gente já tem, são coisas como por exemplo: "É possível fazer um estudo sobre os ônibus e o seu comprimento contratual? ", "É", "Então a gente precisa disso muito rápido". Nesses casos que é muito comum, que eu assumo o projeto, em geral a gestão é minha mesmo e rápida, lista de tarefa, fulano faz isso e me entrega, e vamos seguir, uma coisa muito simples mesmo (Funcionário E1 PENSA sobre o método de implementação de projetos).

As entrevistas indicam, portanto, que havia um método de implementação ágil, embora não de uma maneira oficial. Quando um projeto é de médio ou longo prazo, o PENSA utiliza oficialmente um método de desenvolvimento ágil, como pôde ser captado nas entrevistas.

Agora quando o projeto se torna um projeto mais de médio prazo, estruturado, como por exemplo, essa questão dos painéis de ônibus [...] Então eu preciso que isso seja recorrente, para transformar esse estudo, que foi feito adhoc, em recorrente. A gente passa aí sim para um modelo de gestão de projetos. Então a gente tem a classificação em 4 etapas das tarefas, [...] um modelo ágil e rápido de desenvolvimento e o que está feito a gente joga para cá. Então esse é o modelo que a gente trabalha para esses projetos. Então, quando é um projeto que entra em produção, a gente faz desse jeito (Funcionário E1 do PENSA sobre projetos de longo prazo).

Com efeito, o PENSA usa o método de implementação ágil, principalmente nos maiores projetos. Esses resultados vão de encontro à literatura que destaca que, destaca que, dentre os diversos métodos de implementação, é de se esperar que projetos de *Big Data* são mais adequados para um processo de desenvolvimento ágil e interativo, pois os ciclos de tempo curtos, com resultados rápidos e envolvimento constante do usuário, entregando gradativamente uma solução de negócios facilitam a percepção dos benefícios. De fato, de acordo com Fayet (2015), um método de gestão da implementação tradicional não permitiria um número suficiente de iterações.

Por conseguinte, o presente estudo encontrou evidências de que o método de implementação de projetos ágil é relevante para à implementação e uso de *Big Data*, sobretudo por sua capacidade apresentar resultados rápidos.

### 6.3.3. Abordagem Faseada e Evolucionária

O uso de abordagem faseada e evolucionária, com o objetivo de uma implementação gradual, é constantemente destacado como um fator crítico de sucesso para a adoção de sistemas de informação (CLIFFE, 1999; POON e WAGNER, 2001; FINNEY e CORBETT, 2007; GUPTA, GUPTA e SINGHAL, 2014).

Nesse sentido, é interessante destacar que o PENZA não se distanciou do que é destacado na literatura. De fato, a implementação do novo órgão parece ter ocorrido em fases. No entanto, um aspecto que deve ser destacado na abordagem do PENZA é que ela não foi previamente planejada, nem muito menos explicitada para os quadros internos, conforme pode ser visto nos trechos destacados a seguir.

Sim, nós estamos trabalhando em fases, mas se você me perguntar, essas fases estão planejadas desde o início? Não. Nós planejamos a 1ª fase, sabemos que essa não é uma fase única, mas a gente ainda não tem planejado o que vai ser feito depois, as fases seguintes a gente ainda não tem. Estamos focados nesse problema agora, em melhorar esse ponto e depois verificarmos outras possibilidades e ver quais são as outras fases desse projeto (Funcionário AC3 de uma das áreas clientes a respeito do uso de uma abordagem faseada e evolucionária).

Cara não foi uma coisa explícita. Não (Funcionário E5 sobre como o desenvolvimento em fases não foi planejado).

É interessante destacar, no entanto, que isso ocorreu, em parte, devido à falta de conhecimento sobre como proceder. De acordo com um dos funcionários, o PENZA era, na época, uma iniciativa única no Brasil. Mesmo internacionalmente, eram poucas as iniciativas do uso da análise de grandes conjuntos de dados desestruturados no setor público. Hoje, ainda são poucas as iniciativas conhecidas pelos funcionários do PENZA.

Não foram previstas no projeto. O que estavam fazendo era uma coisa nova, esse grupo PENZA é único no Brasil. Tem um outro pouco semelhante em São Paulo, não é um grupo como o nosso. Não tem o que copiar, nós vamos fazendo e aprendendo, ajustando o caminho. Isso aí é feito em cima de acertos e erros e também é feita em função de demandas pontuais, como acabei de falar da Copa do Mundo ou eleição, isso são demandas pontuais. O PENZA não foi criado em fases [...] Não foram fases, foram mudanças organizacionais naturais (Funcionário E2 do PENZA sobre a abordagem em fases).

Apesar dessa falta de planejamento prévio, foi possível perceber que houve sempre uma preocupação com um desenvolvimento gradual do departamento, o desenvolvimento da estrutura, uma boa aceitação por parte das outras secretarias. Tanto que, inicialmente, o PENZA foi desenvolvido por meio de um projeto piloto, com o uso de análises de dados simples que trouxessem insights interessantes e fáceis de serem compreendidos,

conforme pôde ser observado nas entrevistas de funcionários de áreas clientes e do PENSA.

Assim, já até falei, isso saiu do papel, a gente fez o projeto piloto. O piloto se comprovou o estudo teórico e esse ano estamos implementando isso de uma forma, como operação continuada. Então assim, acho que está tendo todo o desdobramento possível que poderia ter, está tendo (Funcionário AC3 de uma das áreas clientes).

Muitas vezes o que o Pensa faz a gente usa uma massa grande de dados, mas são coisas simples e coisas simples também geram produtos muito bons. Então não necessariamente a gente precisa fazer imagens de cruzamento de dados, regressões múltiplas, bizarras, mas a gente faz o simples usando dados que seriam combinados por outros órgãos, cada um olha ali para o seu e faz muito bem o seu trabalho, mas na hora de juntar, de fazer essa coisa multidisciplinar faltava uma equipe que conseguisse agregar isso tudo e desse coisas mais transversais, resultados mais transversais. Então o *Big Data* como ferramenta e o Pensa pouco antes de eu ter saído a gente usava coisas bem simples. Não tinha um banco de dados nada desse tipo, banco de horas extras era o mesmo... o tecnológico não era esse... Passou a ser agora (Funcionário E5 sobre o início com análises simples).

Então, por exemplo, os meus primeiros momentos, o nosso principal trabalho era conseguir encontrar os dados, era conseguir colocar no servidor, era conseguir estruturar um servidor, um banco de dados, era conseguir trazer tecnologias novas que não fazem parte do cardápio da prefeitura, por exemplo, a questão de dados geográficos, dados de giz. Todas essas tecnologias no começo a gente tinha um trabalho muito grande de identificar, mapear e trazer junto com a estruturação do servidor, junto com a estruturação da coleta de dado, isso eu diria que foi mais ou menos, junto a produção de ideia. A gente tinha que começar a fazer tudo ao mesmo tempo e o primeiro ano foi muito disso de estruturação e etc (Funcionário E1 do PENSA sobre a abordagem faseada).

Acho que tenha acontecido. Estou a pouco tempo, talvez não fale certo. Acho que talvez o PENSA tenha passado por algum momento (antes de eu entrar) de muita estruturação. Precisava montar os bancos de dados, precisava acertar as API's, as conversas com os bancos de dados das outras secretarias, estruturar isso aqui dentro. Depois o PENSA passou por um momento de fazer uns produtos, uns painéis que andamos fazendo com algumas secretarias (Funcionário E3 do PENSA sobre o momento inicial).

De acordo com o representante da Casa Civil, esse início de estruturação usando análises simples foi bastante importante por ter disso o meio pelo qual foi possível constatar que era benéfico investir no desenvolvimento de um departamento de análises de dados tal como o PENSA. É por meio de um sucesso inicial que se garante o patrocínio da alta-gestão, por exemplo.

Como todo projeto, você tem um lançamento inicial no final de 2013, início de 14 onde ele surge na prefeitura, de repente seis malucos que pensam em informação se oferecendo para as frentes de secretarias para disponibilizar informação para a tomada de decisão. Eu acho que o primeiro momento do PENSA foi muito positivo porque ele atendeu diretamente as demandas do prefeito, da alta administração. Então você conseguiu continuar despertando o interesse da alta administração e efetivamente eles conseguiram gerar um

resultado de curto prazo, então esse foi o começo do PENSEA (Representante da Casa Civil a respeito da importância do momento inicial).

Então primeiro é a criação da equipe mesmo que nem existia aqui e depois a seleção das pessoas eu acho que isso é muito importante até para evitar turn over [...] Aí depois tem a parte de integrar todos os dados. Então foi muito importante estar no Centro de operações. Lá é um grande integrador de dados, não de todos, mas de uma grande parte. E aí a gente começou a segmentar e jogar para um banco nosso [...] Então teve essa fase de guardar, de buscar outras fontes. Mas sempre tem coisa chegando então tem que ter essa prospecção e aí depois tem a nossa parte de análise (Funcionário E5 do PENSEA).

De acordo com os entrevistados, somente a partir do sucesso inicial que é possível pensar em fases futuras, novos desafios, institucionalização, conforme pôde ser destacado das entrevistas do representante da Casa Civil e dos funcionários do PENSEA.

Nessa segunda fase, o desafio que eles têm é conseguir fazer, porque o primeiro ano, como tiveram uma série de projetos, muita coisa aconteceu, eles geraram resultados, as secretarias começaram a demandar estudos, e aí o desafio do segundo momento é você ter capacidade de priorizar aqueles estudos que devem ser feitos com base aí em uma matriz de risco-retorno para efetivamente gerar ainda mais resultado para a prefeitura. Então a nossa preocupação nesse segundo momento no PENSEA, é a gente conseguir direcionar o trabalho deles, de forma alguma limitando a capacidade criativa deles, mas tentar dar um pouco mais de foco ao trabalho deles para a gente conseguir ainda mais utilidade nos problemas da prefeitura, então eu diria que é o segundo estágio. Então eles tiveram o lançamento, agora estão na segunda fase que é uma fase de estruturação, de foco, de delivery de determinados serviços e em um terceiro momento, logo logo, a gente precisa começar a pensar na institucionalização dessa turma (Representante da Casa Civil).

A partir agora desse segundo ano é que a gente está com a capacidade de começar a coletar, pegar dados que a gente já tem no banco e começar uns estudos de formas mais rápidos (Funcionário E1 PENSEA sobre a abordagem faseada).

Passamos um tempo estruturando o banco de dados, talvez um tempo trabalhando com demandas internas e agora o PENSEA começa a migrar, ouvir mais a população e entender as demandas externas e trazer isso para dentro da prefeitura (Funcionário E3 do PENSEA).

Por conseguinte, a implementação do *Big Data* parece ter ocorrido de acordo com um processo natural de fases em que primeiro desenvolve-se um projeto piloto para verificar-se a viabilidade e a relação custo-benefício, depois, implementa-se formalmente e, a partir de então, desenvolve-se estudos junto às secretarias. No entanto, tão claro como a ocorrência de uma abordagem faseada, ficou a falta de planejamento dessa abordagem, embora isso não desvalorize a implementação gradual em si. Quando questionado diretamente sobre a importância de uma abordagem faseada, o

representante da Casa Civil – secretaria à qual o PENSA está subordinado – foi enfático sobre a importância.

É fundamental, para cada fase as expectativas são totalmente diferentes, os compromissos são totalmente diferentes e a abordagem é totalmente diferente. Então no começo a gente precisava, todo mundo precisava conhecer os, talvez eles não precisavam escolher o projeto que gerasse um maior resultado, talvez em alguns momentos na largada você tinha que fazer um projeto que gerasse maior visibilidade, para todo mundo você se fazer conhecido e a partir daquele momento você começar efetivamente a construir projetos mais estruturante para a segunda fase. Na terceira fase, eu estou preocupado com institucionalização, eu estou preocupado em como que eu garanto essa política, como que eu garanto que essa cultura de *Big Data* vai ser incorporado dentro da máquina pública (Representante da Casa Civil sobre importância da abordagem faseada).

A resposta obtida de fato está alinhada com a literatura acadêmica ao afirmar que as expectativas são diferentes em fases diferentes. De fato, de acordo com a revisão bibliográfica, a abordagem faseada contribui para o acompanhamento dos resultados obtidos e adaptação às mudanças, bem como contribui para gerir as expectativas com respeito ao sistema de *Big Data*. De acordo com a teoria, “a abordagem faseada revela-se importante por causa da necessidade de adaptação à mudanças, que podem ser decorrentes tanto de forças externas como de forças internas” (INSTITUTE FOR ADVANCED ANALYTICS, 2014).

Apesar do PENSA não parecer ter sido planejado quando discutido como um todo, o mesmo não pode ser dito quando o assunto em questão é o desenvolvimento de estudos e projetos em conjunto com as secretarias. Embora tenha ficado claro que mesmo no relacionamento com as secretarias não há um desenvolvimento de longo-prazo, o desenvolvimento de protótipos e mecanismos de teste pode facilitar a aceitação por parte das demais secretarias e assim aumentar as chances de sucesso da implementação e do uso do *Big Data*. Conforme foi destacado na literatura, a prova de conceito é desenvolvida para ajudar os colaboradores e os usuários a terem noção do resultado que está sendo desenvolvido com a implementação do *Big Data* (PATTERSON, 2014). Patterson (2014) e Leininger (2015), aliás, descrevem passo-a-passo de como obter valor do *Big Data* e desenvolver uma prova de conceito, de modo a maximizar a possibilidade de sucesso do projeto.

O que costumamos fazer muito é antes de falar qualquer coisa, fazer um protótipo aqui dentro. Então pegamos uma amostra dos dados, fazemos análises. Se ficar legal, apresentamos para a secretaria. E já indicando as limitações dos bancos de dados, por exemplo, até onde aquilo representa a realidade para evitar uma expectativa muito grande. Costumamos gerar expectativa no momento em que temos algum produto para entregar, é difícil



ficar falando muito. Normalmente vamos as reuniões e perguntamos quais os dados, podemos apresentar algumas ideias, mas bem vago para não deixar nada ‘amarrado’. O interessante de fazer um piloto e apresentá-lo para as secretarias é porque ali já emerge algumas críticas e para corrigir erros (Funcionário E3 do PENSA sobre a prototipação).

Quando começamos o trabalho com o PENSA não existia um planejamento inicial de que atividades iam ser desenvolvidas. (Hoje) não existe porque a gente não tem como montar um planejamento estratégico como o Pensa à longo prazo. Para eu montar um planejamento estratégico com fases eu tinha que ter um acordo de nível de serviço com o PENSA para entender e eu não consigo isso, porque eu não consigo ter uma demanda calculada [...] Ele é demandado por todo mundo e hierarquicamente quem manda mais demanda mais, tem as prioridades e eu não consigo fazer isso. Então quando a gente tem a oportunidade de fazer, “Pensa, podemos continuar?” “Podemos”. “Então nós vamos fazer mais um pedacinho aqui” e assim vai, não seria um grande planejamento (Funcionário AC2 de uma das áreas clientes sobre impossibilidade abordagem faseada na parceria SMTr-PENSA).

Dessa maneira, embora não tenha sido verificado um planejamento prévio de uma abordagem faseada, foi possível identificar uma consciência por parte dos entrevistados a respeito de um desenvolvimento faseado. As expectativas dos atores envolvidos com a implementação mudaram ao longo do tempo. Os desafios também. Além disso, foi possível perceber que o uso de uma abordagem por meio de protótipos junto às secretarias – um dos motivos que faria da abordagem faseada um fator relevante à implementação e uso do *Big Data* – pode ser verificada por meio das entrevistas.

Por conseguinte, tendo os achados na revisão bibliográfica e os resultados constatados por meio das entrevistas, proponho que a abordagem faseada é um fator relevante na adoção e uso do *Big Data*. Embora a abordagem adotada não tenha sido planejada, verificou-se consciência da importância de uma implementação gradual. Isto posto, proponho o seguinte:

<b>Proposição 2</b>	<b>Três fatores relacionados à implementação são relevantes para adoção e uso de Big Data no setor público: o uso de estruturas facilitadoras da implementação; o uso de métodos de implementação ágil; e o uso de uma abordagem faseada e evolucionária.</b>
2.1	O uso de pelo menos uma estrutura facilitadora à implementação – comitê gestor, líder/campeão de projetos e grupos de superusuários – é relevante na adoção e uso de <i>Big Data</i> .
2.2	O método de implementação de projetos ágil é relevante para à implementação e uso de <i>Big Data</i> , sobretudo por sua capacidade apresentar resultados rápidos, que aumenta a percepção de valor dos usuários.
2.3	A abordagem faseada é um fator relevante na adoção e uso do <i>Big Data</i> , tendo em vista que ela favorece uma implementação gradual e por meio de protótipos junto aos demais departamentos da organização, contribuindo para a gestão das expectativas e dos desafios, que mudam ao longo do tempo.

## 6.4. FATORES RELACIONADOS À PERCEPÇÃO DE VALOR DE NEGÓCIOS

### 6.4.1. Gestão de Expectativas

As expectativas em relação ao *Big Data* são muito elevadas, o que pode favorecer ao desapontamento. Por conseguinte, é importante considerar que existirão dificuldades de performance na utilização das novas tecnologias, e que é necessário gerir as expectativas das potencialidades da tecnologia e as chances de sucesso. Recomenda-se, por exemplo, um longo período de experimentação que inclua testes independentes daqueles realizados por indivíduos que vão operar o sistema, bem como envolvimento dos responsáveis pela implementação do *Big Data* nas discussões dentro da comunidade científica e acadêmica (FREEMAN, 2013).

De fato, de acordo com um dos funcionários de uma das áreas clientes, as expectativas com relação ao PENSA eram muito altas dentro da prefeitura.

Acho que no 1º momento da existência do PENSA houve sim esse bum de "ah, o PENSA pode fazer tudo, pode fazer qualquer tipo de análise", talvez até pedidos esdrúxulos. Não estou falando de um caso concreto. É uma proliferação de pedidos, acredito que lá dentro tivesse coisas assim não tão importantes, sem um filtro, mas essa coisa foi se assentando com o tempo (Funcionário AC3 sobre as expectativas acerca do PENSA).

Conforme um dos funcionários comentou em sua entrevista, no entanto, as expectativas eram altas junto àqueles que conheciam o *Big Data* e o PENSA, o que significa que não eram todos.

Em relação a Casa Civil que é o chefe mesmo, a Casa Civil tinha uma expectativa alta. Em relação a outras secretarias ao longo do tempo teve algumas que foram com o trabalho que a gente foi desenvolvendo junto foram aumentando o seu nível de expectativa. Outras a gente nunca conversou, nunca teve nada, provavelmente nem conhece a gente só viu na revista o que é o Pensa, mas nunca trabalhou junto. Então a expectativa é zero porque não tem conhecimento dele (Funcionário E5 sobre a gestão de expectativa).

Isso traz a intuição de que não houve um programa de Gestão de Expectativas propriamente dito, e sim que a expectativa ia se ajustando à realidade de acordo com o desenvolvimento dos estudos junto às secretarias e a obtenção de resultados. No entanto, o funcionário da Casa Civil deixou a entender o gerenciamento das expectativas era uma atribuição do gestor do PENSA.

A gestão de expectativas é um trabalho quase que diário do gestor em relação a esse projeto, porque ele recebe a diretriz da alta administração, ele desdobra

essa diretriz para o dia a dia dele, então ele tem uma orientação, ele tem um foco, a prefeitura trabalha com meta, o gestor trabalha com planejamento e ele sabe exatamente qual é a diretriz da administração. E ao mesmo tempo ele tem demandas de todas as secretarias toda semana querendo fazer estudos diferentes. Então eu acho que não tem foco não, é um trabalho quase que diário dele para gerenciar tudo isso. O que a gente tenta fazer é dentro da diretriz do prefeito ele tem uma folga, e essa folga dele, ele vai gerenciar as expectativas dos clientes dele, sejam clientes da administração, sejam as outras secretarias (Funcionário da Casa Civil).

De acordo com a entrevista de um dos representantes de uma das áreas clientes, no entanto, a gestão de expectativas não parece ser tão importante para a implementação e uso do *Big Data*, tendo em vista que os trabalhos desenvolvidos em conjunto com o PENSA são pontuais.

Então aqui não existe nenhuma área da prefeitura que possui pessoal de tecnologia. O serviço é prestado pelo IPLAN e agora o IPLAN, que não tinha ninguém de *Big Data*, usa o Pensa como uma equipe de *Big Data*. Então como nós não somos os contratantes, nós não temos esse planejamento estratégico com o Pensa de desenvolver projetos à longo prazo. A gente tem coisas bem instantâneas, bem pontuais que vão acontecendo e o Pensa vai nos ajudando e é demandado por todas as secretarias. O Pensa hoje é demandado por todo mundo, por todas as secretarias (Funcionário AC1 de uma das áreas clientes).

Esses resultados encontrados vão no sentido contrário ao que foi disposto na revisão bibliográfica. É interessante, entretanto, ressaltar que a experimentação foi uma forma de se lidar com a criação de expectativas.

O que costumamos fazer muito é antes de falar qualquer coisa, fazer um protótipo aqui dentro. Então pegamos uma amostra dos dados, fazemos análises. Se ficar legal, apresentamos para a secretaria. E já indicando as limitações dos bancos de dados, por exemplo, até onde aquilo representa a realidade para evitar uma expectativa muito grande. Costumamos gerar expectativa no momento em que temos algum produto para entregar, é difícil ficar falando muito. Normalmente vamos as reuniões e perguntamos quais os dados, podemos apresentar algumas ideias, mas bem vago para não deixar nada ‘amarrado’. O interessante de fazer um piloto e apresentá-lo para as secretarias é porque ali já emerge algumas críticas e para corrigir erros (Funcionário E3 do PENSA).

Com o tempo, as expectativas se ajustaram sem que houvesse um desapontamento, conforme havia sido sugerido por Freeman (2013). Isso pôde ser observado na entrevista de um dos funcionários das áreas clientes, conforme trecho a seguir.

Hoje a gente já tem, a prefeitura já tem uma capacidade de avaliar melhor o que o PENSA pode e o que o PENSA não pode, o que faz sentido para ele, o que não faz e o que eles vão te dar de retorno. Acho que no início houve muito uma expectativa de que o PENSA fosse resolver o problema e não é isso que o PENSA faz. O PENSA ajuda a trazer respostas para aquele problema para que você ataque os pontos críticos com maior assertividade, mas não é o PENSA que vai resolver o problema (Funcionário AC3 de uma das áreas clientes).

Como não foi possível verificar um programa de Gestão de Expectativas propriamente dito, não é possível verificar se o mesmo é, de fato, um fator relevante para a implementação e uso com sucesso do *Big Data*. O que foi verificado é que a experimentação foi uma forma eficiente de se lidar com a questão da gestão da expectativa.

#### **6.4.2. Comunicação**

Outro aspecto que se destacou como candidato a fator relevante na adoção e uso de sistemas de análise de dados anteriores ao *Big Data* foi a comunicação. De acordo com o levantamento teórico, a comunicação revelou-se relevante desde a promoção formal das equipes de projeto, até a comunicação do escopo, a educação a respeito dos novos processos de negócios, a cooperação entre os diversos departamentos da empresa e a resolução de conflitos, conforme pode ser observado na revisão bibliográfica.

De fato, as entrevistas revelaram que a comunicação é um desafio de maneira geral tendo em vista que o PENSA depende da colaboração das secretarias e às vezes do Estado para fornecer o dado correto e acurado, conforme revelaram funcionários da equipe em entrevista.

Mas nós encontramos bastante resistência quando chega. Primeiro como somos muitos ‘pequenos’ (comparado com a prefeitura), falta interação com as secretarias. Às vezes desenvolvemos uma ferramenta interessante, e a secretaria não sabe ou não tem desejo de usar. Outras vezes também desenvolvemos outras ferramentas que não são tão úteis, mas é porque não estamos vivendo o dia-a-dia da secretaria (Funcionário E5 do PENSA).

A Polícia Militar, por exemplo, não registra acidentes. Então a prefeitura tem que se comunicar com o estado, tem que pegar o dado do estado, esse dado é fornecido à prefeitura, mas é fornecido a um órgão da prefeitura em cima de um documento de sigilo daquele dado. Então o PENSA está querendo utilizar uma base de acidentes e não consegue. Porque base de dados da guarda municipal é muito boa, mas eu não posso tirar conclusões precisas daquilo dali porque ela não é a única que registra acidente (Funcionário E2 do PENSA).

Já aconteceu algumas vezes de projetos que não saíram por que a secretaria não queria que saísse. Coisas bobas. Por exemplo, a secretaria achava que tinha muito trânsito e não deveria comunicar aquilo (Funcionário E4 do PENSA).

No fim as contas eles são excelentes analistas, excelentes técnicos, mas não tem a vivência do dia a dia, então não adianta você produzir um dado que a gente não vai usar. Tipo, para que eu quero esse dado? (O que você produziu) eu já sei. Então assim, eu entendo que teve um momento de desconfiança inicial, porque é um órgão novo, está ignorando a experiência da gente? Não, desconfiança inicial, mas entendo que isso pode ser facilmente ultrapassado, desde que sejamos demandantes, que é o que está acontecendo agora (Funcionário AC2 de uma das áreas clientes).

Embora seja vista como um desafio, a comunicação se revelou importante em diversos pontos para o PENSA. Como exemplo, pode-se citar a importância da comunicação para a compreensão inicial do que era o PENSA e a qual objetivo ele se propunha. Além disso, a comunicação revelou-se importante também para não haver necessidade de recorrer ao decreto que permitia ao PENSA ter acesso a quaisquer dados pertencentes a quaisquer secretarias e departamentos da prefeitura. Esses exemplos ficam claros nos trechos de entrevistas destacados a seguir.

Eu acho que o relacionamento do PENSA com as secretarias, primeiro ele é importante para que as secretárias entendam o que o PENSA faz, entendam a capacidade e as possibilidades do que o PENSA pode prestar de serviço e como consequência possam lembrar do PENSA e solicitar a ajuda do PENSA em determinados problemas em que se não houver essa, esse relacionamento mais próximo talvez não lembrasse da existência do PENSA e não vislumbrasse como que o PENSA pode ajudar na resolução daquele problema específico. Então assim eu acho que é muito importante o PENSA ter relacionamento próximo com os órgãos. O relacionamento interno (Funcionário AC3 de uma das áreas clientes).

Externamente é muito importante o relacionamento com as secretarias porque muitas vezes equipes como essas, isso também eu vejo com o IPP que é o órgão de dados da prefeitura tem que estar muito bem alinhado para não achar que você está se intrometendo num assunto que não é seu. Então pensando meio que na casa civil então mostrando mais trabalho foi ganhando mais facilidade de penetração nas secretarias. Antigamente tinha muita dificuldade de conseguir os dados assim, eu sempre pedia. Até saiu um decreto depois que dá acesso a todos os dados da prefeitura, a gente nunca precisou desse decreto para nada assim. Nós sempre tivemos bastante uma vontade de ajudar. Até politicamente assim tem um trabalho. A gente não percebe, mas tem um trabalho político, para que a gente está trabalhando com aquilo e eles vão se beneficiar também (Funcionário E5 do PENSA).

Dentre os diversos fins para os quais a comunicação auxilia a implementação e uso de uma tecnologia, destaca-se a resolução de conflitos internos. Para tal, é necessário haver uma comunicação eficiente com todos os envolvidos no projeto e todos aqueles que serão impactados pelo projeto. Isso pôde ser verificado nas entrevistas, conforme trecho destacado na entrevista de um dos funcionários do PENSA.

Tem casos que assim o conflito é natural, acontece, por exemplo, se a gente for falar de um sistema simulador de engarrafamento que a gente fez, que mostra como o engarrafamento se espalha pela cidade, quais são os pontos críticos quando vai se vai fazer um fechamento, isso é importante para as obras. Aí quando a gente passa esse *insight* é claro que em certa medida acaba se sobrepondo ou conflitando um pouco com estudos que os órgãos já fazem, já estavam fazendo. Inicialmente, a primeira reação que é natural "Espera aí, a gente já faz isso, a gente já tem alguma coisa que vai nessa linha", aí é o momento de fazer, da gente sentar e ver "se para vocês é feito de uma forma, a gente achou uma outra forma de fazer, vamos ver se a gente consegue juntar essas duas e produzir uma terceira forma que agregue o melhor de cada um desses modelos" (Funcionário E1 do PENSA sobre a ocorrência de conflitos e exemplificando a comunicação com um meio de resolução de conflitos).

Nesse sentido, é interessante notar que a comunicação favorece a compreensão das capacidades do sistema e contribui para revelar as expectativas, além de reduzir a resistência das secretarias e conquistar o apoio das mesmas – o que no fim reduz a ocorrência de conflitos. A importância da comunicação com esse intuito também pôde ser verificada nas entrevistas, principalmente como uma responsabilidade do cargo de gestor da equipe.

É uma pessoa que não só integra a equipe, mas também faz a gestão das expectativas, é uma pessoa que atua junto às outras áreas para garantir que nós não somos x9, nós só estamos aqui para ajudar, vocês tem muito a ganhar com a gente e a gente tem muito a ganhar com vocês (representante da Casa Civil).

A gente não sabe nem que potencial aquilo tem. Então assim, como a gente não sabe que potencial aquilo tem, a gente não demanda nada, a gente fica ali observando e ali eles produzem o negócio e a gente fica "poxa, mas isso eu já sei". Quer dizer, acho que há uma falta de informação, mas é uma falta de preparo e a gente no decorrer do processo foi descobrindo isso. "Ah, eles fazem isso, poxa, eles pegam informação em tempo real? Eles visualizam isso? Eles têm acesso a essa informação?". A gente pode elaborar isso e demandar deles (Funcionário AC2 de uma das áreas clientes).

É importante, no entanto, destacar que embora a comunicação tenha sido estimulada por um lado, por outro, a alta gestão requisitou que os integrantes do PENSA agissem com liberdade. Tanto que a validação dos *insights* obtidos pelos estudos por vezes era realizada por meio de comunicação direta com a área cliente, mas por vezes não. Um dos intentos da alta gestão era que o PENSA trouxesse inovação na solução dos problemas da cidade. Por conseguinte, foi solicitado que fossem sugeridas ideias novas e não fossem executadas apenas ideias sugeridas pelas secretarias, como pôde ser observado pela entrevista de um funcionário do PENSA.

Isso é uma coisa que alguns sim, outros não, mas que poderia ser melhor acompanhado ou ficar mais próximo da secretaria ou mais próximo daqui mesmo. As vezes tinham uma ideia genial, mas falavam “nada ver” A gente “Tá bom, você conhece, respeito, lógico”. O prefeito falou: Eu não quero que vocês se contaminem com os órgãos. Tipo não é para resolver o que eles precisam. É dar uma ideia do que eles estão pensando em fazer, do que eles já tentam. Então se eu sou um cara que faz sempre a mesma coisa eu não vou perceber que eu posso mudar um processo aqui. Se você chega de fora e olha você vai conseguir mudar. Então a gente ficando um pouco mais distante a gente tem possibilidade de fazer isso, mas por outro lado a gente não conhece muito do assunto e essa é uma dificuldade (Funcionário E5 do PENSA).

Além disso, a comunicação contínua estimula a troca de informações entre os usuários e a equipe de implementação do sistema. A literatura também recomenda que a comunicação deve ocorrer entre todos os *stakeholders* e deve informar o progresso do projeto.

Nesse contexto, vale lembrar ainda que a contínua cooperação entre os departamentos e a administração de problemas são essenciais para o sucesso dos projetos de *Big Data*. Os conhecimentos adquiridos a partir de um projeto de análise de *Big Data* podem contribuir para a melhora do desempenho de diferentes outros projetos e departamentos, incluindo as equipes de TI, cientistas de dados e analistas, e as pessoas em toda a organização a partir da linha de gerentes de negócios. Linhas de comunicação abertas são necessários para manter os projetos no caminho certo e para entregar valor de negócio (INSTITUTE FOR ADVANCED ANALYTICS, 2014).

Nas entrevistas, não foi verificada uma realidade distante dessa, tendo em vista que foi por meio da comunicação que foi possível desenvolver projetos que realmente interessavam às secretarias, conforme pode ser visto na entrevista de um dos funcionários do PENSA.

Isso é normal em qualquer desenvolvimento novo. A gente tem que adaptar isso para as necessidades daquele órgão e o órgão vai passando para a gente nas necessidades, isso é super comum (Funcionário E1 do PENSA).

Por fim, vale destacar que a literatura indica que a coleta de informações a partir de cada projeto ou iteração e posterior comunicação das lições aprendidas contribui para um aumento das probabilidades de sucesso de projetos futuros. Embora não seja algo institucionalizado, é interessante verificar que há uma prática internalizada, pois os mesmos analistas costumam desenvolver projetos nas mesmas áreas continuamente, conforme pôde ser observado nos trechos de entrevistas a seguir.

O que existe é o seguinte, se concentrou mais em um técnico do Pensa nos projetos de setor de transporte e com certeza esse cara já tem mais conhecimentos da área de transporte, inclusive é a pessoa que está sendo alocada para a gente. Então ele hoje é a pessoa mais capacitada à entender e trazer um resultado mais rápido. Qualquer outro ia fazer um negócio “não bem assim” demoraria um pouquinho mais o desenvolvimento e ele talvez tivesse um pouco mais de capacidade de velocidade, não é nem capacidade. Capacidade de fazer uma coisa mais rápida (Funcionário AC1 de uma das áreas clientes).

Com efeito, o presente estudo encontrou evidências confirmatórias em relação ao que foi levantado na revisão bibliográfica. Em outras palavras, quando bem executada, a comunicação pode aumentar as chances de sucesso da implementação e, portanto, é um fator relevante na implementação e uso do *Big Data*.

### **6.4.3. Monitoramento e Avaliação de Performance**

A literatura acadêmica afirma que *Big Data* não foge à norma dos demais sistemas de informação. Aferir o sucesso é um processo difícil. É por isso que é importante

construir objetivos, medidas e metas mensuráveis e monitoráveis que demonstrem os benefícios de uma equipe focada em *Big Data Analytics* (OHLHORST, 2013).

Nesse sentido, foi interessante constatar pelas entrevistas que não foi possível perceber uma diretriz de acompanhamento de resultados na implementação do *Big Data* e do PENSA, e sim uma percepção dos resultados dos estudos realizados pelo departamento ao analisar determinados dados. Isto é, o que se percebe não é uma avaliação do desempenho do PENSA, mas como seus estudos alimentam a cultura de gestão por indicadores e o aprimoramento da tomada de decisão na Prefeitura.

É, houve assim, a mudança foi muito radical. Não foi um trabalho do Pensa que motivou, mas o trabalho do Pensa nos ajudou a tomar uma decisão que era tirar todo o transporte público de dentro de Copacabana. Foi isso que nós fizemos no Réveillon. Estamos no 2º Réveillon, fizemos isso também na Jornada Mundial da Juventude, onde não entra transporte público em Copacabana, ônibus, porque o metrô está lá, não tem como tirar. Não entra ônibus em Copacabana durante o período do evento. O que a gente precisava era colocar esses ônibus em pontos estratégicos com saídas regulares para determinadas áreas da cidade. Então nós medimos o tempo de deslocamento. Foi muito mais rápido porque os ônibus não enfrentavam mais aqueles engarrafamentos enormes dentro de Copacabana. Por outro lado muitas pessoas se acham penalizadas por terem que caminhar, elas têm que caminhar, elas têm que sair de Copacabana e ir a pé para pegar a sua condução. Mas quando chegava era bem organizado em filas, as pessoas entravam nos ônibus e rapidamente estavam nos seus destinos (Funcionário AC2 de uma das áreas clientes).

Então era nosso interesse também fazer com que a guarda municipal tivesse a sua satisfação aumentada. Essa meta específica que eu estou me referindo é meta de satisfação. Mas como é que a gente consegue fazer com que a satisfação suba? Como é que a gente estratifica a insatisfação do cidadão para poder encontrar os problemas na operação e corrigi-los. A gente tinha uma dificuldade aqui e recorremos ao PENSA e pedimos a eles para fazer um estudo num caso que gerou um mapa de calor em que foi apontado nesse mapa uma ponderação entre as áreas da cidade com maior demanda e maior insatisfação. Então nós chegamos em pontos críticos da cidade, onde tinha uma demanda muito grande por estacionamento irregular, fiscalização de estacionamento irregular e uma baixíssima situação. Em conjunto com isso pegamos os dados das pesquisas de satisfação e analisamos os motivos da insatisfação, tudo isso junto com o PENSA e chegamos a conclusão que a maior insatisfação era proveniente de um motivo de recorrência da infração. Então esse estudo nos fez chegar a conclusão de que nós precisávamos de uma fiscalização ostensiva nesses lugares, nessas áreas da cidade. Como é que isso aumenta a satisfação? Na verdade isso não aumenta diretamente a satisfação, mas ela diminui a quantidade de chamados de pessoas insatisfeitas. Então você não aumenta direto a satisfação, mas você aumenta a insatisfação porque você tira pessoas insatisfeitas do seu universo de análise. Essa foi a nossa primeira experiência com o PENSA que foi muito proveitosa, nós fizemos alguns pilotos ano passado, nós fizemos apenas em áreas pilotos e esse ano nós estamos fazendo ainda como operação continuada. Mas nas áreas pilotos ano passado em algumas delas nós tivemos um aumento de quase 50 pontos percentuais na satisfação naquela área (Funcionário AC3 de uma das áreas clientes).



De acordo com outros entrevistados, no entanto, não há um acompanhamento e monitoramento de resultados. Pelas entrevistas deles, ficou a impressão de que a avaliação de resultados é um desafio, seja pela falta de acompanhamento ou pela falta de feedback, conforme pode ser visto pelos trechos destacados a seguir.

Então é uma coisa que eu acho que é um desafio que é esse feedback que a gente pensa. Esse feedback é uma coisa que precisa melhorar (Funcionário E5 do PENSA).

Isso é um problema [...] a gente não tem a cultura de avaliar. Eu acho que é um processo, mas a gente ainda não faz isso, o que tem de interessante ali no trabalho do Pensa, é que você tem coisas muito estruturais, e você tem medidas de muito curto prazo, para medida de curto prazo você consegue avaliar o impacto com muita facilidade, para isso a gente avalia, a gente consegue avaliar o impacto financeiro, avaliar o risco, porque elas são mais simples, mas efetivamente uma cultura de você, fez a análise, implementou, qual é o impacto disso no curto, médio e longo prazo, fez sentido, não fez sentido, vamos redirecionar a política. A gente ainda tem que aprender, é um trabalho de amadurecimento (Entrevistado que pediu para não ser identificado).

Ainda não fiz nenhuma avaliação posterior de uma implantação de projeto, não sei qual foi o impacto de fato ou não, mas também a gente acaba trabalhando muito com o produtor de conhecimento e aí não tem nenhum impacto direto, por exemplo, nesse estudo do BRS, analisamos quais linhas estão piorando mais ao longo do tempo em termos de velocidade média. A gente nem faz uma proposta para isso, só fazemos o estudo e enviamos para a Secretaria, não tem como avaliar o impacto disso. Na verdade o problema é quase de matemática, então, eles não tem matemática suficiente no mundo, a estatística não é suficiente para inferir se foi impacto da proposta ou se foi algum outro motivo (Funcionário E3 do PENSA).

Eu assim, não sei te dar essa resposta com total certeza, eu enquanto cliente do PENSA não fiz, não participei dessa análise, não sei te dizer se o PENSA internamente fez essa análise (Funcionário AC3 de uma das áreas clientes).

Apesar de não haver encontrado nenhuma evidência a respeito do acompanhamento e monitoramento da performance, não há dúvidas de que o PENSA pode se beneficiar se alguma iniciativa do tipo for institucionalizada. De acordo com a revisão bibliográfica, o acompanhamento da implementação deve estar atento aos custos de implementação, a qualidade dos resultados obtidos, bem como deve permitir a comparação entre o antes e o depois da adoção do *Big Data*. De fato, se houver um acompanhamento nessas dimensões, fica muito mais fácil diminuir eventuais resistências com relação à nova tecnologia e perceber, suas potencialidades e seus benefícios.

Essa situação, no entanto, parece que será modificada no futuro próximo. De acordo com um dos entrevistados, a partir desse ano o PENSA terá metas compartilhadas com o restante da prefeitura.

Esse ano, o gestor está trabalhando metas compartilhadas com a secretaria, que seria uma terceira maneira de ‘botar óleo nessas estruturas’. Num momento que eles tiverem a mesma meta que nós, todo mundo vai fazer acontecer. Até ano passado, o PENSA tinha metas separadas (metas próprias), então eles poderiam cumprir nossa meta, que era fazer o estudo, mas outra meta legal que era complementar não era cumprido por não ser meta da secretaria. Ter uma meta só para todos (Funcionário E3 do PENSA).

Em suma, ao contrário do que consta na literatura acadêmica, não foi possível verificar evidências, por meio das entrevistas, de que o monitoramento e acompanhamento de resultados é um fator relevante na implementação e uso de *Big Data*. O sucesso dos projetos é aferido mais por uma percepção qualitativa e isto pode estar relacionado com a questão da cultura da experimentação.

#### 6.4.4. Ética

Por meio da literatura acadêmica foi possível verificar que a questão da ética e do uso dos dados tem sido bastante discutida quando *Big Data*. Muitos dos debates têm sido a respeito de questões epistemológicas, tais como quais dados podem ser utilizados em uma análise. Outras questões tais como denominação de responsáveis por avaliação de questões éticas, sobre que condições alguém deve ser considerado como parte de um grande conjunto de dados e o que significa alguém ser identificado ou ser analisado sem saber também têm sido discutidas (BOYD e CRAWFORD, 2012; TENE e POLONETSKY, 2013). É possível perceber que a utilização ética do *Big Data* tem sua relevância bem discutida na literatura acadêmica.

Devido aos seus objetivos, o PENSA não possui contato direto com a sociedade. Há uma maior preocupação com a ética quando está se lidando diretamente com dados dos cidadãos, conforme revelado nos trechos de entrevistas exemplificados ao largo desta seção.

É, nós ainda não passamos nenhum problema de questão de ética com os dados, primeiro porque nós não temos... Nada que o PENSA faz vai para mídia, porque PENSA não faz para a mídia direto, para os cidadãos. O PENSA faz é para o órgão. Então nessa questão de ética como você está colocando, entre o PENSA e o próprio órgão: isso é resolvido internamente (Funcionário E2 do PENSA).

A estrutura da maneira como foi configurada é importante não apenas porque o PENSA não possui contato com a sociedade, mas também porque ela permite uma verificação dupla da ética relacionada à qualquer questão ou dado que possa vir a ser analisado, como dito pelo representante da Casa Civil.

Esse ponto é muito importante, porque você tem, enfim, a filosofia do ponto de vista ético, das pessoas que conceberam. Havia uma preocupação muito grande da alta administração de se respeitar toda a privacidade da informação. Então havia uma diretriz, da alta administração, para não haver qualquer possibilidade de você usar essa informação de maneira antiética. [...] Então o que eu estou querendo dizer para você é o seguinte, a gente para manter fase de determinados estudos, o gestor eventualmente pode precisar de dados fazendários, de dados de... Da receita da fazenda, dados de IPTU, por exemplo. Esses dados de IPTU, no banco de dados da fazenda você consegue nominalmente pelo CPF, saber quanto que a pessoa está pagando eventualmente o IPTU, isso é um dado fiscal extremamente sigiloso... Extremamente sensível, que ninguém pode ter acesso, só a pessoa da fazenda que tem acesso a esse banco de dados, dificilmente, eu acho pouco provável, que a secretária da fazenda, ao repassar esses dados para o PENSA, ela não tem a preocupação de não passar dados sigilosos. Então o que eu estou querendo dizer, ela não vai deixar de passar o dado do IPTU, mas não vai passar por pessoa, a área de *Big Data* não terá a possibilidade de analisar [...]. Então é como se fosse um Double check, entendeu? É tipo assim, você tem a diretriz de não faça de cima para baixo, e você tem o movimento de baixo para cima que eu não deixo você fazer, eu não te dou chance de fazer. Ainda assim, eventualmente eles podem ter uma informação sigilosa que a gente entende que, enfim, que eles vão respeitar as diretrizes, e não vão manipular, fazer mau uso dessa informação, mas é um dado muito interessante, por exemplo, os dados de saúde são muito sensíveis do ponto de vista do sigilo. Dificilmente o Pensa vai trabalhar com uma base de dados que você tenha um dado de saúde por pessoa, não pega (Funcionário da Casa Civil).

De fato, durante as entrevistas foi possível perceber que tanto as áreas clientes como os próprios indivíduos do PENSA são responsáveis por fazer uma análise ética das questões. Primeiro, há uma análise na própria área cliente. Se, por ventura, o dado é liberado para análise, então o funcionário do PENSA pode refletir se pode, de fato, utilizar aquele dado da maneira que ele é disponibilizado ou não e tem o poder de se recusar a usar determinado conjunto de dados.

Uma grande preocupação aí, isso é muito específico do 1746 porque nós tratamos, lidamos com dados de cidadão. Temos por um lado os dados que ele fornece para a execução do serviço, para relatar problemas da cidade, que na verdade é isso, quando ele pede um serviço ele está relatando um problema na cidade, mas junto a isso vem dados pessoais daquele cidadão que fez aquela solicitação. Nossa grande preocupação é jamais deixar que esses dados saiam desse ambiente que é o 1746, mesmo que assim, para o PENSA que é um co-irmão nosso, isso a gente não pode fazer, somos muito criteriosos com relação a isso (Funcionário AC3 de uma das áreas clientes a respeito da ética no uso dos dados).

Só o secretário de transporte é a pessoa que poderia determinar alguma coisa disso, mesmo assim discutindo com o prefeito (Funcionário AC1 de uma área clientes a respeito da ética no uso dos dados).

Felizmente todos os contatos que a gente tem com empresas, se mostraram sempre absolutamente éticos e muito cuidadosos nessa questão. Então uma empresa vai passar um dado para a gente sempre desanonimizado, é uma preocupação que está muito à vista (Funcionário E1 do PENSA a respeito do uso dos dados).

Se não me engano os dados que são *top secrets*, esse acho que nem nós temos acesso, e são alguns dados que nós temos acesso, o nosso acesso é limitado (Funcionário E2 do PENSA).

Não. Eu acho que não. Vai da formação de cada um ali dentro mesmo e do caráter, do que cada um pensa que é ética (Funcionário E5 do PENSA a respeito do uso dos dados).

Esse mecanismo em mão dupla verificado nas entrevistas é importante porque analistas de *Big Data* raramente reconhecem que há uma diferença considerável entre estar em público e de ser público (BOYD e MARWICK, 2011). Só porque o conteúdo é acessível ao público não significa que ele foi concebido para ser utilizado por qualquer pessoa. Há questões sérias envolvidas na ética da coleta e análise de dados on-line (ESS, 2002). O processo de avaliação da ética em uma pesquisa não pode ser ignorado simplesmente porque os dados são aparentemente públicos.

Por conseguinte, essa questão é de máxima importância até porque os usuários não são necessariamente conscientes de todos os usos múltiplos e outros ganhos que vêm de informações que eles postaram. E de fato, uma vez que os dados são dispostos para o PENSA, se crê que eles possam ser analisados e, portanto, a análise é feita. Até mesmo porque, conforme foi dito por um dos funcionários do PENSA, a equipe não tem interesse nos dados individualizados.

Cara eu acho que tudo que tinha lá era para usar. Isso é mais responsabilidade dos órgãos que coletam os dados do que nossa. Então se eles disponibilizaram aquilo é porque pode usar (Funcionário E5 do PENSA).

A literatura ainda levanta a questão de que deve haver discussões no sentido de impor limites éticos para o que será feito com o *Big Data*. Faz-se necessário, por exemplo, discutir critérios de análise bem como responsáveis por análise para caso de dúvidas futuras à respeito do uso (ou não) de determinado dado. Tais preocupações puderam ser verificadas também dentro da prefeitura, conforme trechos destacados a seguir.

Eu sou esse perfil aqui por ter a formação jurídica, por ter conhecimento nessa área, sendo a minha área de atuação no direito, eu mesmo tomo essa decisão, mas, como eu te disse, eu faço questão de que já nem venha, e o único caso que a gente tem isso, para ser muito honesto, na administração pública, isso é fácil, vamos dizer que eu estou em uma situação bastante confortável aqui com relação a isso porque dentro da administração 99.9999% dos dados são públicos. Então eu tenho como paradigma anterior a decisão do projeto, mas que quase em todos os casos isso não é importante porque os dados já são públicos. A única questão na administração pública que tem um sigilo mesmo de dados, são as questões fiscais por uma questão constitucional. Então se for para fazer qualquer estudo com esses dados, a gente tem que pensar muito bem, é o único dado dentro da administração. E claro assim, dados pessoais eventualmente têm cadastro de funcionário, isso são coisas que a gente nem entra ali [...] Agora, para os dados de terceiros,

todos os terceiros que a gente trata, essa preocupação acaba não se mostrando necessária por conta de os próprios parceiros já tem essa preocupação. Claro que no longo prazo isso pode vir a se tornar um problema, mas não acho que... Pode ser que apareçam questões problemáticas no médio, longo prazo, com relação a isso, mas se for da forma como a gente está fazendo aqui, isso não vai... (Funcionário E1 do PENSA).

Eu acho que tem a primeira instância que é a do gestor, e ele é responsável pelo órgão, responsável pelo projeto, e ele vai fazer uma primeira avaliação dele: "eu acho que a gente pode usar essa informação ou não dessa maneira", se ele tiver dúvida, ele vai subir para a casa civil, a segunda instância, se a gente tiver dúvida, vai subir para o prefeito, e esse é o fluxo normal da coisa (Funcionário E1 do PENSA).

Não tem uma pessoa que cuida disso do tipo: "olha primeiro o dado passo por mim e depois vai para vocês" Tem base que a gente consegue direto com o órgão (Funcionário E2 do PENSA).

Tendo em vista a análise das entrevistas, foram encontradas evidências de preocupação com as questões éticas que me levam a crer que, em algum grau, a ética é um fator relevante à implementação e uso de *Big Data*.

#### **6.4.5. Privacidade e Segurança**

Conforme pode ser visto na revisão bibliográfica, privacidade e segurança cibernética são provavelmente as principais preocupações para aplicativos de *Big Data*. Privacidade refere-se principalmente a informações pessoalmente identificáveis (PII), isto é, informações que podem ser usadas para identificar um indivíduo.

A questão do anonimato, por exemplo, é uma que tem suscitado muitas discussões na academia. Minelli, Chambers e Dhiraj (2013) acreditam que os dados recolhidos para um fim específico podem ser anonimados e depois ser utilizados para outros fins como a identificação de padrões coletivos, sobretudo quando não há necessidade de voltar aos dados individuais. O desafio, segundo o autor, é que quanto mais anônimos forem os dados, menor sua utilidade.

Para tornar ainda mais complexa a discussão, em 2006, um grupo de pesquisa com sede em Harvard conseguiu tirar do anonimato partes de um conjunto de dados supostamente anônimos (BOYD e CRAWFORD, 2012). Esta percepção de inexistência do anonimato se refletiu nas entrevistas.

Não, não tem anonimato. Tudo você pode voltar atrás se você se esforçar (Funcionário E3 da equipe do PENSA).

A questão do anonimato, no entanto, é apenas uma das muitas questões relacionada à privacidade do dado. Na literatura acadêmica, por exemplo, foi possível observar que

alguns estudos defendem que a noção de privacidade pode ser alterada por fatores contextuais tais como geolocalização e diferenças culturais (MINELLI, CHAMBERS e DHIRAJ, 2013; YAN, 2013).

Na opinião do gestor do PENSA, a questão da privacidade e do anonimato antecede até mesmo as três preocupações propostas por Yan (2013) - (a) quais informações pessoais podem ser compartilhadas e com quem, (b) se as mensagens podem ser trocadas sem que ninguém mais vê-los, e (c) “se” e “como” alguém pode enviar mensagens de forma anônima –, conforme pode ser visto pelo trecho destacado a seguir.

A proteção de dados pessoais não visa a proteção do indivíduo. Apesar de se falar em dados pessoais, você está visando a proteção da liberdade de uma sociedade, você está tratando da questão de como, é aquilo que se diz, se a todo momento alguém soubesse tudo o que você está fazendo, você não teria nunca liberdade para atuar espontaneamente, você estaria sempre se policiando e isso é horrível, isso tolhe a evolução, isso cerceia a criatividade inclusive. Então essa visão ditatorial do uso de dados é algo que me preocupa radicalmente e é por isso que, por exemplo, eu não acho necessário que a pessoa para colaborar com o dado, precise ter o dado identificado, você pode colaborar muito com dados absolutamente desidentificados. Tem a questão da constituição que é vedada ao anonimato, mas dependendo do tipo de coisa que você está fazendo, eu não vejo problema [...] A gente tem essa preocupação pelo cidadão, porque me parece que a sociedade ainda não está pronta, ainda não, não é que não está pronta, ela ainda não amadureceu, não refletiu o suficiente para saber a importância que tem esses dados pessoais. Então, no papel de órgão público aí como indutor de política, a gente tem essa obrigação de assumir uma indução desse respeito a privacidade e aos dados pessoais (Funcionário E1 do PENSA).

Essa afirmação demonstra a existência de uma preocupação com a privacidade dos indivíduos que geraram os dados até certo ponto. Por um lado, conforme pode ser visto na análise sobre ética, há uma preocupação acerca da ética na análise dos dados, a tal ponto que existe um canal de verificação dupla: tanto a área proprietária do dado como os funcionários do PENSA são instruídos a refletir sobre a ética no momento de efetuar uma análise. Por outro, no entanto, não há uma crença que de fato a privacidade total seja algo possível, conforme pôde ser observado.

Essa situação, no entanto, não foi considerada tão preocupante pelos entrevistados. Um dos funcionários do PENSA destacou, por exemplo, que os dados quando chegam ao PENSA, já chegam autorizados para a análise. Ou chegam viam um acordo que impede a divulgação do dado.

(Quando os dados chegam) eles já não estão na base. Quando é crítico, sim, a gente tem alguns outros problemas de dados que são de empresas que não podemos fornecer para fora. O *Waze*, por exemplo, é um deles, o *Waze* vende os dados dele, para gente ele fornece como parceria. [...] Quando vem um dado muito crítico interno da prefeitura, ele já vem normalmente autorizado.

Ou no caso dos alunos, vem o nome, escola aonde eles moram, não é um problema tão grande mas ainda assim, esse dado não pode sair, fica aqui dentro (Funcionário E3 do PENSA).

O anonimato do dado favorece a privacidade do usuário:

A gente hoje não tem nem condição de acessar esse tipo de informação, a Riocard tem esse dado, uma vez que ela tem o cartão que é nominal e você tem por onde ele andou você sabe quem andou. A secretaria de transporte não recebe esse dado nesse formato, ele já recebe o dado agrupado por viagem, eu não sei quem andou, eu sei que pessoas andaram (Funcionário AC2 de uma das áreas clientes).

Por causa de questões como essa, na academia há um certo consenso de que a garantia de transparência sobre os métodos de coleta e análise de dados é a melhor maneira de lidar com questões de sensibilidade. Simplesmente permitir que os indivíduos controlem o que "funciona para eles" e forneçam loops de feedback adequados é a melhor postura a ser adotada.

Embora não tenha transparência a respeito de quais dados utiliza em suas análises, o PENSA, como instituição, é dotado da preocupação relacionada à transparência para os cidadãos, sobretudo quando está em debate a questão dos dados sensíveis. Não só o PENSA como a prefeitura do Rio como um todo, e isso pôde ser percebido não só na entrevista com um dos funcionários do PENSA – conforme trecho destacado a seguir –, como também no fato da prefeitura adotar toda uma postura de transparência de seus dados – conforme os demais trechos que foram destacados.

Claro que têm alguns dados que por questões legais são um pouco mais problemáticos, por exemplo, os dados tributários que envolvam dados pessoais, aí é uma questão constitucional inclusive de sigilo dos dados. Então há dados sensíveis que a gente tem que trabalhar com um cuidado maior e aí os órgãos tem uma resistência que não é uma resistência pessoal. Então esse dado aqui eu não posso passar porque ele é muito sigiloso, dado de contribuinte, o quanto o cara pagou, onde pagou nota eletrônica e etc (Funcionário E1 do PENSA).

Até onde eu entendo, essa disponibilização de dados de terceiros não ocorre, o Pensa ele tem acesso dados a base de terceiros, de algumas bases. Essas bases, se o terceiro autoriza que seja disponibilizado, é disponibilizado, se o terceiro não autorizar, ele não disponibiliza, isso dentro de uma cultura, você deve ter visto lá, de open data, a gente está tentando o máximo possível trabalhar com (...book), dar o máximo de transparência possível para as pessoas trabalharem (Representante da Casa Civil).

Terceiros em termos da prefeitura muitos assim. Uma das coisas foi o Portal Data Rio tinha o que podia, o que não podia...acesso aos bancos. Então tem acessos externos da prefeitura e internos de outros órgãos também (Funcionário E5 do PENSA).

Olha, até hoje, não compartilhamos dados com terceiros, a minha preocupação que eu estava falando é com o terceiro externo, nem pensar. O

que é compartilhado com externo é o que nós temos no site de transparência, se você entrar lá é transparente para todo mundo, aquele dado que está lá é compartilhado. Para eu passar esse dado para outro, esse outro tem que entrar em contato com o órgão, eu não tenho autorização nem por decreto nem por nada de passar o dado, o estudo é independente (Funcionário E2 do PENSA).

Isto é, o posicionamento da prefeitura é tal que mesmo o compartilhamento de dados com terceiros não é feito, por mais que o PENSA seja um departamento que preste serviços apenas internamente. Conforme destacado, o único compartilhamento de dados é àquele feito via canal oficial da prefeitura, canal este que toda a sociedade tem acesso: o Portal Data Rio.

A privacidade não é o único problema associado com *Big Data*. A segurança é outro. Conforme destacado na revisão bibliográfica, *Big Data* traz grande risco ao garantir grande acúmulo de informações em um único local. A garantia de segurança dos dados é importante para diminuir a resistência à implementação das tecnologias associadas ao *Big Data*. Nesse sentido, um dos entrevistados destacou que há um sistema de segurança baseada no controle da permissão de acesso.

Só a gente tem acesso ao banco, quando o dado sai é ou pela plataforma de (dado aberto), ou então alguém vem pedir [...] Há uma questão de login e senha (Funcionário E3 do PENSA).

Isso, no entanto, não é suficiente para avaliar a segurança como fator relevante para a implementação e uso de *Big Data*, tendo em vista que somente um entrevistado comentou algo a respeito e que as possibilidades de invasão deste sistema não foram discutidas.

Dessa maneira, o presente estudo encontrou evidências de que a privacidade, tal qual a ética, é um fator que vale a pena ser investigado mais profundamente, tendo em vista que ele se encontra no cerne da discussão acerca do *Big Data* e, sobretudo, que algumas das oportunidades e benefícios estão associados justamente à resposta para perguntas como “que tipo de dados podem ser utilizados na análise?”. O mesmo, no entanto, não pode ser dito da segurança: embora haja evidências na referência bibliográfica a respeito de sua relevância para implementação e uso do *Big Data*, não houve discussão suficiente para encontrar evidências à respeito de sua relevância.

Com efeito, proponho o seguinte:



Proposição 3	Dos fatores relacionados à percepção de valor do <i>Big Data</i> , destacam-se como fatores relevantes a experimentação, a comunicação, a ética e a privacidade.
3.1	As expectativas dos atores envolvidos variam ao longo da implementação e uso do <i>Big Data</i> . No setor público, a gestão das expectativas dá lugar a experimentação dos diferentes tipos de análise possíveis, na busca de insights, é um fator de geração de valor na adoção e no uso de <i>Big Data</i> .
3.2	A comunicação é um fator relevante na implementação e uso do <i>Big Data</i> . A comunicação deve ocorrer desde a promoção do projeto até a resolução de conflitos internos, pois contribui para a compreensão de todas as capacidades do sistema. A comunicação contínua estimula a troca de informações e deve ocorrer entre todos os stakeholders e deve informar o progresso do projeto.
3.3	O monitoramento e a avaliação de resultados não é um fator de percepção de valor relevante na implementação e uso do <i>Big Data</i> no setor público no Brasil. O uso de mecanismos de implementação pode contribuir para garantir a eficácia da implementação, mas não implica em uma percepção de valor tal que o monitoramento em si seja relevante na implementação, pois os resultados podem ser avaliados qualitativamente.
3.4	A ética é um fator de percepção de valor relevante na implementação e uso de <i>Big Data</i> no setor público no Brasil. Um planejamento prévio deve ser feitos de maneira a impor limites éticos, discutir quais dados podem ser analisados e sob que circunstâncias. Devem ser apontados os responsáveis por decidir sobre questões éticas que não forem previamente analisadas.
3.5	O respeito à Privacidade é um fator de percepção de valor relevante na implementação e uso de <i>Big Data</i> no setor público no Brasil. A garantia de transparência sobre os métodos de coleta e análise de dados é a melhor maneira de lidar com questões sensíveis. A Segurança, por outro lado, não é um fator de percepção de valor relevantes considerando o mesmo contexto.

## 6.5. NOVOS CANDIDATOS À FATORES RELEVANTES

Durante as entrevistas alguns desafios e fatores relevantes que não haviam sido antevistos na revisão bibliográfica foram apontados. Nesse sentido, destacam-se, sobretudo, a percepção dos entrevistados a respeito da estrutura, da infraestrutura e dos desafios relacionados aos dados e sua gestão propriamente ditos. Vou discutir cada um desses tópicos separadamente.

### 6.5.1. Estrutura

Conforme pode ser visto anteriormente, o estabelecimento de uma estrutura centralizada estava associado intimamente aos objetivos do PENZA de atuação transversal, melhora na troca de dados e informações entre as secretarias, e, principalmente, estímulo à

tomada de decisões baseadas nos dados. Esse tipo de estrutura foi usado com sucesso na implementação do MODA, o modelo no qual o PENSA se baseou.

De fato, a princípio essa assimilação obteve bastante sucesso. Mesmo a escolha do espaço físico para localizar a equipe foi bem-sucedida, devido ao fluxo de informações do Centro de Operações do Rio (COR), conforme pode ser visto no trecho a seguir destacado da entrevista de um dos funcionários do PENSA.

O fato de estar aqui no centro de operações é muito importante porque a gente vê o que está acontecendo na cidade o tempo todo, tem dado para coletar, para analisar, isso é extremamente positivo, uma lição. E o centro de operações na sua própria natureza, é muito mais ágil, alheio as burocracias por conta da sua emergência de ação e isso para gente é uma lição que não tem preço, como as coisas funcionam aqui (Funcionário E1 do PENSA).

Até a disposição do PENSA como um departamento à parte e conexão direta junto à Casa Civil, tal qual ocorreu com o MODA e a prefeitura de Nova York, deve ser destacada como importante, tendo em vista que o apoio e patrocínio do *top management* são importantes para o sucesso de uma iniciativa de implementação de análise de dados, conforme pode ser visto anteriormente. Nesse sentido, a escolha foi tão bem-sucedida que, na visão do representante da Casa Civil, ela não precisa ser nem mesmo repensada.

Eu acho que em algum momento talvez, ainda não está maduro o suficiente para ser repensada, é aquilo que ele falou, fazer a capacitação, disseminar a cultura para dentro da máquina, isso tudo são coisas a serem trabalhadas. Agora, mudar a estrutura do Pensa hoje, eu acho que não está na hora, precisa passar mais três fases aí antes (Representante da Casa Civil).

No entanto, essa visão não foi confirmada pelos funcionários e representantes das áreas clientes. De uma maneira ou de outra, eles acabaram revelando que o ideal seria haver uma conexão mais permanente entre o PENSA e suas áreas clientes, talvez tal qual uma estrutura federada, em que há uma unidade central, mas há também unidades menores que atuariam em conjunto com as secretarias permanentemente. Isso pode ser observado nos trechos destacados à seguir.

Outras vezes também desenvolvemos outras ferramentas que não são tão úteis, mas é porque não estamos vivendo o dia-a-dia da secretaria. O ideal seria se tivesse pequenos PENSAS dentro de cada secretaria, seria melhor utilizado. O potencial é gigante, mas talvez não conseguimos fazer tão bem por estarmos afastados da prefeitura. Seria importante manter uma unidade central, mas talvez essa seria mais uma unidade de dados, porque é importante ter isso para a cidade. Talvez um banco de dados com tudo o que a cidade tem. [...] (Então seria mais uma estrutura de federação? Você tem pequenas PENSAS que respondem a uma unidade central) Seria superinteressante chegar na secretaria de educação e ter um grupo sentado lá e conversar nossa língua com eles. Por que normalmente nas reuniões vem uma pessoa que entende mais da secretaria, mas que mesmo assim

conversamos em uma outra linguagem (Funcionário E3 do PENSEA sobre a estrutura).

Eu queria um Pensa só para mim (Funcionário AC1 de uma das áreas clientes).

Vejo assim que a secretaria tem que se profissionalizar para ter profissionais do tipo do PENSEA aqui. Então assim, é inadmissível a gente não trabalhar com *Big Data* [...] Acho que eles já nos mostraram uma demanda, isso é muito bom. Aí é isso que estou falando, a gente não quer mais trabalhar (num estudo isolado com o PENSEA), a gente gostaria de ter profissionais, a gente precisa de profissionais com esse perfil dentro da secretaria (Funcionário AC2 de uma das áreas clientes).

Pelas entrevistas, foi possível perceber a ausência, nas áreas clientes, de um profissional com perfil analítico com competências relacionadas a dados. Nesse sentido, é, portanto, importante destacar que a estrutura em si não é um desafio, e sim que há um potencial de novos e maiores ganhos que podem – e devem – motivar uma reflexão a respeito da estrutura do PENSEA ou dos processos de seleção de funcionários para as demais secretarias.

Talvez o PENSEA não esteja fazendo algo que o MODA implementou em seus processos: a automatização das soluções e dos estudos propostos às secretarias. Não foi investigado se isto tem sido feito ou não, mas certamente isto seria uma alternativa para ampliar o uso de análise de dados nas secretarias. A conclusão é de que, dado o contexto no qual está inserido a prefeitura do Rio de Janeiro, há evidências de que a estrutura do departamento de análises é um fator relevante na implementação e uso do *Big Data*. Com efeito, a estrutura deve ser constantemente repensada para aumentar a captação de benefícios oriundos da análise de dados, e seu planejamento deve considerar o contexto e a cultura. A centralização da análise de dados em uma área pode, a longo prazo, ser revista para que permita que os cientistas de dados estejam cada vez mais próximos e melhor compreendam os processos das áreas clientes, em modelos mais descentralizados.

### **6.5.2. Infraestrutura/Arquitetura de Sistema**

Além das questões relacionadas à estrutura, outro assunto que chamou a atenção nas entrevistas foi a infraestrutura de sistema. Não só pelas dificuldades que surgiram, mas também por deixar bem claro que as dificuldades relacionadas à infraestrutura dependem bastante do objetivo da organização.

Embora a literatura tenha levantado a questão de que *Big Data* está mais relacionado com a falta de estrutura dos dados e com a velocidade na qual os dados são captados e

analisados do que com o gigantesco volume de dados propriamente dito, as entrevistas evidenciaram que tudo é uma questão do que é valorizado pela organização que adota à tecnologia. No caso da prefeitura do Rio de Janeiro, tanto a análise em tempo real como o uso de dados não estruturados não foram grandes dificuldades, tendo em vista que não eram o maior interesse da prefeitura, conforme trechos destacados a seguir.

Olha a gente, o uso de dados não estruturados ainda é muito novo, e muitas vezes para ser honesto para você, o custo de trabalho de análise em troca do benefício que ele traz ainda hoje, não sei se tem uma boa relação em relação ao que a gente consegue. Eu acabo usando o dado não estruturado mais como uma fonte para as análises dos meta-dados estruturados que ele me fornece. O uso do dado não estruturado, como eu disse, ainda é muito, para as nossas necessidades, e para o que a gente pode fazer com os dados estruturados hoje ainda, eu acho que é um horizonte que está um pouquinho distante, que eu acho que precisa ser mais desenvolvido (Funcionário E1 sobre o uso de dado desestruturado).

A maioria das questões não, a maioria das questões pode ser feita com prazo, a maior parte dos casos sim, a questão é que a menor parte dos casos que precisam de dados em tempo real, em geral são os mais críticos. Então você tem esse (trade off) de respostas rápidas para casos críticos, e repostas não tão rápidas para casos não tão críticos [...] (Além disso) se você tem uma queda, por exemplo, de uma captura de um determinado dado 24 horas, 48 horas, é crítico, para aquele momento, é crítico, mas quando você olha, você tem 1 ano de dado daquilo, 1 hora, 24 horas eventualmente sem o dado, é prejudicial suficiente para o estudo naquele momento, mas quando você está trabalhando com *Big Data*, pegando um volume grande de dados, que você está mais interessado nas tendências e nos (nortes) e não no individual, essas questões não são totalmente críticas. (Funcionário E1 do PENSA sobre o uso de análise em tempo real).

Agora aquela coisa que eu gostaria, eu vou ler placa? Vou. Eu consigo ler placa em tempo real? Pode ser que sim. Se eu conseguir ler a placa em tempo real eu posso solicitar ao estado a relação de carros roubados? Posso. Se eu leio a placa em tempo real eu posso checar para verificar se acabou de passar um carro roubado? É tempo real? É. Então eu não posso mandar uma mensagem para a PM dizendo que um carro roubado acabou de passar aqui? Isso funciona, ter dado real de leitura de placa para pegar, agora, outra coisa, é uma demanda da prefeitura? Não. Segurança pública é do Estado. Pode ser que o Estado solicite nossos dados para que ele desenvolva esse projeto. Então esse é o mínimo de coisas em tempo real (Funcionário E2 do PENSA).

Apesar de não haver necessidade de análise em tempo real, ocorreram problemas relacionados ao tempo de processamento de modelos, que se estendia por muito mais tempo que era adequado, conforme destacado a seguir.

Ainda esbarramos bastante em capacidade de processamento, é um problema bem forte até hoje. Muita coisa não vai dar para fazer... A máquina deve ser uma promessa (...) inclusive a gente tentou analisar os dados aqui na máquina atual por 10 dias sem sucesso, desistiu de tentar. A gente rodou em 50 minutos na máquina da EMC, a máquina que foi recentemente adquirida. Acho que vai mudar bastante isso (Funcionário E3 do PENSA sobre problemas relacionados ao tempo de processamento).

Os funcionários ainda afirmaram que os desafios relacionados ao tempo de processamento não eram únicos. Havia também problemas relacionados à capacidade computacional e o armazenamento de dados.

Já, aí a gente tenta sair por outras soluções, seja por utilização de amostras, seja por utilização de outras bases, seja por utilização de outras tecnologias, como, por exemplo, as vezes um negócio não roda em um ambiente que a gente está rodando, a gente tenta fazer isso com o pré-processamento reduzindo esses dados (Funcionário E1 do PENSA).

Já teve, foi o que eu acabei de falar para você, essa mudança de arquitetura vem sempre de um tempo maior de processamento. Não só dentro de processamento, mas também armazenamento, espaço, vai trocando de servidor, vai aumentando a capacidade dele. Então a estrutura como um todo foi pensada para se você tiver menor tempo de processamento e mais espaço para os dados e isso aí foi modificando realmente ao longo como você está colocando (Funcionário E2 do PENSA).

Alguns problemas bem contundentes de velocidade de ônibus, de carro, *Waze* e tal a gente desistiu de atacar porque viu que não tinha capacidade computacional (Funcionário E3 do PENSA sobre a capacidade)

Sim. Sim. Sempre caía. Inicialmente a gente pedia maior capacidade e foi considerando o máximo que podia dar para a gente. E aí depois quando lotava tinha que apagar coisas que a gente não estava usando mesmo. E também de instabilidade. [...] Tinha problema de cair, problema de infraestrutura. A conexão tinha redundância de rede, de perder dados [...] (Além disso) conforme foi aumentando o número do volume de dados foi ficando complicado e de vez em quando estagnado aí tinha que zerar o negócio... Não zerar, mas apagar algumas coisas que não estavam sendo usadas porque já não cabia mais ou então o processamento não conseguia entregar a tempo, eles demoravam muito para rodar algumas coisas (Funcionário E5 do PENSA sobre problemas relacionados ao processamento e à confiabilidade da infraestrutura de sistema).

Além disso, foram identificados alguns desafios relacionados à confiabilidade das máquinas, como quedas de energia, falta de capacidade de armazenamento, ou simplesmente o uso de uma máquina com confiabilidade menor para poder realizar determinada captação de dado, conforme um dos funcionários deixou bem explícito.

Para o Twitter, tivemos problema, bastante problema para conseguir salvar dentro do banco de dados, dentro do servidor, dentro da prefeitura. Não consigo rodar meu script que salva Twitter direto do servidor. Então preciso rodar no meu local e aí uma vez por semana cai porque falta luz. Então barramos, mas por problema de proxy, não precisa zerar como problema de infraestrutura, talvez sim, algumas coisas não conseguimos rodar por causa disso, acho que só. O servidor é confiável em termos de acesso a internet, mas como ele tem um proxy, não posso usar ele. Aí preciso usar meu local que é pouco confiável (Funcionário E3 do PENSA).

É porque não existe um planejamento de futuro, mas eu acho que arquitetura, eu acho que não é toda ruim não, a maneira que está utilizando é que não está muito boa [...] se for querer guardar dados, tudo bem, mas faz uma manutenção de dados, guarda em um arquivo morto dados antigos, a gente não precisa ter dados de 2 anos atrás no banco de dados fresco, agora, pelo menos parte entre servidores diferentes, se quiser juntar depois, faz um novo [...] E depois que o banco de dados fica ruim, tudo dá errado, até as coisas triviais. “Quando é que vai chegar o novo banco de dados”, porque na

verdade não tem nenhuma aplicação fluente usada de verdade atualmente, tipo se o banco de dados cair durante 1 mês, as pessoas não vão perceber (Funcionário E3 do PENSA sobre problemas relacionados ao processamento e à confiabilidade da infraestrutura de sistema).

Um dos funcionários chamou a atenção de que o problema relacionado a falta de confiabilidade é mínimo, tendo em vista a existência de mais de uma cópia dos dados:

Não porque o (gestor) é muito consciente dessa parte, a gente trabalha com cópias. Hoje nós temos aqui na nossa arquitetura, esse aqui é o banco que vai receber inserções, esse gera uma cópia aqui aonde eu acesso essa cópia (Funcionário E2 do PENSA quando questionado sobre os problemas de confiabilidade).

É importante destacar, no entanto, que, sendo mínimo ou não, desafios relacionados à infraestrutura existiram. Logo, há evidências de que o planejamento e escolha da infraestrutura são fatores relevantes na adoção e uso do *Big Data*.

### **DESAFIOS RELACIONADOS AOS DADOS**

Embora a sociedade viva um momento de sobrecarga de dados, surgiram das entrevistas desafios relacionados diretamente aos dados. Em específico, destacaram-se questões relacionadas à existência de um determinado dado para análise, à qualidade dos dados existentes e ao armazenamento da proveniência dos dados.

Os desafios relacionados à existência dos dados foram percebidos como conhecidos em todos os níveis da administração. O funcionário da Casa Civil, por exemplo, demonstrou conhecer a questão da inexistência do dado, mas afirmou que o PENSA estava conseguindo realizar os estudos apesar desse desafio. Um dos funcionários do PENSA, por outro lado, exemplificou o problema e qual foi a solução encontrada.

Um sim, outras não, mas eles conseguiram conduzir estudos nesses sentidos, eles fizeram um estudo bem interessante também, da dengue, eles usavam mídia social... (Funcionário da Casa Civil a respeito de problemas relacionados à existência de dados para análise).

Começamos a fazer o levantamento dos dados para poder cruzar e falar “olha com probabilidade de x% e um erro de tanto as próximas árvores vão cair nessa região aqui”. Para chegar a isso, tivemos que levantar corredores de ventos. Quem tem essa informação de corredores de ventos? Não conseguimos os dados dos corredores de ventos. Qual foi o outro? Histórico de queda de árvore. Quem tem histórico de queda de árvore? Existe? Existe, aí você vai ver quem tem? [...] não passava na cabeça deles que realmente eles precisavam ter um histórico detalhado de queda de árvore da cidade e que aquilo ali poderia ser feito. Quem mais? Bombeiro também tem. Entramos em contato com um bombeiro. Um oficial deu uma parte, depois quando eu entrei em contato aí já não davam mais. Então dado do estado já não vinha para cá. O que mais? Dados de volume de vento, dado de...um dado que nós trabalhamos bem foi um dado dos pluviômetros. Então de chuva conseguimos, de vento não conseguimos [...] Bom resumindo contato,

praticamente nada porque não tínhamos dados [...] Com isso chegamos à conclusão que tinha que ser desenvolvido senso, com uma ronda na cidade. Porque quem cuida da árvore não sabe que existe a ronda [...] Então essas são as dificuldades, não tem cadastro, os cadastros não se comunicam (Funcionário E2 sobre o desafio da inexistência de dados).

E com isso é legal. Foi criado uma PR, ponto de referencia para contratar uma empresa para fazer processo arbóreo na cidade. E esse censo agora foi desenvolvido também outro programa e nos demos um monte de pitaco o tempo todo nesse programa. E o que nós fizemos? Pegamos todos os atores na época, botamos sentados todo mundo na mesma mesa. “Olha, tem que ser feito isso”. Porque a nossa demanda nós não conseguimos resolver por causa disso, disso e disso, por causa de vocês, esses atores que estão aqui. Então daí saiu esse termo de referência para que seja feito na cidade do Rio de Janeiro um censo arbóreo. Aí não tem jeito, o sistema é um só. Vamos plantar uma mudinha? Essa mudinha vai ganhar um (pedigree), um numerozinho e esse é o mesmo sistema (do responsável pela poda). Então quando você ligar pedindo uma poda de árvore, você antes vai entrar no sistema para verificar qual é a previsão de poda de árvore, você vai ter uma previsão para cada árvore em função desse censo (Funcionário E2 sobre a solução dada ao desafio da inexistência de dados no caso do estudo de queda de árvores).

O fato de ter sido encontrada proposta uma solução, no entanto, não significa que o desafio da não existência de dados deva ser desconsiderado. O desafio surgiu e dada a cultura de não guardar dados, ele deve se repetir em diversos contextos. Principalmente porque esse desafio está intimamente relacionado ao desafio da estrutura, que também surgiu das entrevistas sem ter sido observado na literatura acadêmica anteriormente.

Outro desafio relacionado ao dado que surgiu das entrevistas foi a qualidade do dado. Intimamente relacionado à questão da existência, o problema da qualidade, no entanto, difere pela possibilidade de, com os devidos cuidados metodológicos, ser superado a um custo de tempo e esforço.

Então, aí entra essa questão, por exemplo, a gente tem informação que não é confiável pela sua origem, pela origem e pela natureza, vamos dizer assim, por exemplo, os dados de *Twitter*, quão ele é confiável para uma questão de distribuição geoespacial, você não sabe, você não sabe exatamente quais pessoas, quantos por cento da população usam, etc. Então nesse caso a gente faz alguns trabalhos antes [...] para ver se batia essa distribuição, e bateu, bairros com menor densidade, tem menos pessoas twittando, bairros com maior densidade tem mais pessoas twittando em uma lat-long. Então assim, o dado do *Twitter* se mostrou um dado confiável. Então na questão da qualidade do dado quando ele vem de uma fonte não confiável por natureza, um dado gerado socialmente e etc, e não gerado pela academia, cientificamente, a gente tem esse cuidado de fazer os trabalhos para tentar medir a sua confiança.

Depois você tem o dado que tem que vir com qualidade e nem sempre vem, isso também é muito comum em volumes grandes, o GPS de ônibus com falhas por áreas de sombra, por erro de preenchimento no número da linha no aparelho do GPS, essas coisas acontecem, e um dos trabalhos que a gente fez foi, por exemplo, identificar essas falhas para identificar qual o nosso erro no resultado. Então aí a gente faz a, pelo menos no caso do GPS, a gente fez a identificação de quantos por cento dos carros estão circulando, ou sem preencher o GPS, ou enviando menos o sinal do GPS que deveria por causa de áreas de sombra e etc, um sinal de celular, e isso a gente passa para o

órgão responsável cuidar da melhoria do dado. Então, por exemplo, a (SMTR) está fazendo um trabalho agora grande exatamente disso, melhorar a qualidade do GPS. Então eles estão atuando e fiscalizando as empresas com relação a isso, e por fim tem a qualidade do resultado dos nossos trabalhos, e esse é um trabalho um pouco mais de médio, longo prazo, que eu acho que ainda não obtém o suficiente para a gente medir, porque a maior parte das ideias que a gente passa, como eu te falei, tem um tempo para maturação dentro do órgão, para maturação burocrática, para ser implementado, e aí é que a gente vai conseguir tem um benchmarking suficiente para a gente saber o que precisa fazer (Funcionário E1 do PENSA).

Mas acontece, por exemplo, de ter dado de muito má qualidade e não temos como saber se o dado vem de lá de dentro mesmo, não sei, de alguma forma a secretaria não manda o melhor dado que ela tem [...] O cara tem uma tabela do Excel muito mal feita e a gente precisa trabalhar, (diferenciar, para conseguir estruturar esse dado de maneira legível) e de volume pesado tem problema claro, mas até escrever o script, ficar apanhando para escrever o script que ele normalmente já pré-processada. Quando ele vai para o nosso banco de dados ele já está estruturado. A gente apanha um pouco para conseguir escrever o conector (Funcionário E3 do PENSA).

Sim. A gente também fazia essa sugestão ao dono daquele dado ou ao gerador daquele dado para ver se ele melhora ou descobre qual é o problema que está gerando aquilo (Funcionário E5 do PENSA).

Tal qual a questão relacionada à existência dos dados, o desafio da qualidade pode não ser primordial e ter como ser superado. No entanto, ambos são desafios que devem ser levados em consideração em um planejamento prévio para saber qual estrutura adotar, por qual departamento (no caso da prefeitura, secretaria) iniciar a implementação do *Big Data* e como gerenciar as expectativas sobre os resultados que podem ser obtidos. Dessa maneira, esses desafios parecem ser fatores relevantes na adoção e uso da tecnologia à medida que impactam outros desafios, aumentando sua complexidade, e prejudicando a performance da tecnologia e percepção dos benefícios oriundos delas.

Por fim, outro tópico que se destacou nas entrevistas foi a questão do armazenamento da proveniência dos dados. Embora um dos funcionários tenha afirmado categoricamente que há esse tipo de armazenamento de metadados – o funcionário E2 do PENSA –, entrevistas com outros funcionários esclareceram que o armazenamento é feito de modo incipiente, conforme pode ser visto nos trechos destacados a seguir.

A gente não tem uma estrutura de metadados, mas isso entra como nome das tabelas (Funcionário E5 do PENSA).

De maneira estruturada, bem pouco. As nossas tabelas costumamos nomear os dados começando por onde venham. Se temos escola de educação, Alunos Escolas [...] (Não tem muito uma base de metadado então?) Não, até eu entrar tinha, depois tinha o wiki onde ficava mais ou menos indicado, de onde vem, quem foi o técnico e tal. Caiu em desuso (Funcionário E3 do PENSA).

E é justamente essa maneira incipiente com a qual é feito que pode gerar desafios futuros, caso os dados sejam necessários para novas análises ou verificação dos



resultados encontrados em uma análise já realizada. Sem um armazenamento da proveniência dos dados, se os dados forem perdidos, sua origem não pode ser recuperada e uma análise pode tornar-se invalidada ou ter de ser refeita desde o início.

(E se vocês precisarem, vocês acham que conseguem ir na secretaria?) De uma maneira geral, a gente vai perguntar aqui dentro quem foi o responsável por fazer a integração, mas já tivemos casos da gente perder dados, dado do Strava, que é um aplicativo de bicicleta, que conseguimos, eles vendem, mas conseguimos uma amostra, em algum momento de ciclista que passa no rio, que usa o Strava. A pessoa que foi responsável por receber esse dado no momento saiu do PENSA, está em outro lugar agora. Formatamos a máquina dele, perdemos os dados, e acabou, o dado sumiu. Hoje eles não dão mais dados para a gente, já não temos mais Strava (Funcionário E3 do PENSA).

As entrevistas revelaram, portanto, que existem desafios relacionados aos dados que devem ser avaliados como fatores relevantes na adoção e uso de *Big Data*. Há de se avaliar, no entanto, até que ponto o contexto e a cultura nos quais o PENSA está inserido podem impactar tornar esses candidatos a fatores relevantes como específicos à própria experiência do departamento. Com base nos resultados encontrados, proponho o seguinte.

<b>Proposição 4</b>	<b>Existem fatores tecnológicos e/ou associados à tecnologia que são relevantes na implementação e uso do <i>Big Data</i>. Entre eles se destacam fatores associados à estrutura da equipe de análise de dados em relação à organização, a infraestrutura de sistemas de análise de dados e os desafios relacionados à obtenção dos dados para análise.</b>
4.1	A estrutura do departamento de análises é um fator relevante na implementação e uso do <i>Big Data</i> . A estrutura deve ser constantemente repensada para aumentar a captação de benefícios oriundos da análise de dados, e seu planejamento deve considerar o contexto e a cultura.
4.2	Há evidências de que o planejamento e escolha da infraestrutura de análise de dados são fatores relevantes na adoção e uso do <i>Big Data</i> . O uso de softwares livres e <i>open source</i> e de hardwares capazes de atender às necessidades de análises de dados favorece a percepção de valor.
4.3	A disponibilidade dos dados e de registros históricos, bem como a completude dos dados e uma boa estrutura de metadados são fatores relevantes na implementação e uso do <i>Big Data</i> . A cultura de geração de dados completos é importante para que haja análises de qualidade.

## 7. CONCLUSÃO

*Big Data* não é tanto sobre tecnologia quanto é sobre a transformação de negócios (SCHMARZO, 2013). Discutir *Big Data*, suas oportunidades de negócios e seus desafios correlacionados não é uma discussão sobre um conjunto de tecnologias, e sim sobre a mudança de paradigma que o *Big Data* representa para a sociedade e por conseguinte para as organizações. Se antes não era possível analisar completamente todo um conjunto de dados, hoje o é (BLOEM, VAN DOORN, *et al.*, 2012); se antes não era possível analisar de maneira custo-efetiva dados com estrutura diferente, agora o é; se antes não era possível analisar em tempo real ou quase real, hoje o é (DAVENPORT, 2014). Essas mudanças aumentam a importância dos dados e das informações oriundas de suas análises no ambiente e nos processos de negócios.

É inevitável a comparação entre *Big Data* e os sistemas tradicionais de análise, inclusive porque *Big Data* é, em alguns aspectos, uma extensão das tecnologias de *Business Intelligence and Analytics* (BI&A) (CHEN, CHIANG e STOREY, 2012). O que essa extensão permite fazer, no entanto, vai muito além dos sistemas tradicionais de BI, como foi discutido nesta pesquisa. Apesar de similaridades entre as técnicas analíticas, para sobre as organizações interessadas em aderir às novas tecnologias de *Big Data* e desenvolver projetos *data-driven* a dúvida: os fatores relevantes na adoção e uso de *Big Data* são os mesmos fatores que se revelaram relevantes na adoção e uso de sistemas de análise de dados anteriores, tais como sistemas de *Business Intelligence*?

Para responder a esse questionamento, foi importante primeiro delimitar o que se entende por *Big Data*. Nesta pesquisa, *Big Data* foi definido como o conjunto de tecnologias capazes de armazenar, limpar, tratar, processar, analisar e extrair valor de conjuntos de dados que são difíceis de processar com tecnologias tradicionais (e que podem ser definidos por um ou mais V's do conjunto multidimensional de V's), bem como o próprio conjunto de dados que necessitam que demandam essas tecnologias para serem analisados.

De maneira geral, diversos são os benefícios que podem motivar as organizações a aderir a esses conjuntos de tecnologias e aos vastos conjuntos de dados que tais tecnologias permitem analisar. Não por acaso, as oportunidades criadas pelo *Big Data* estão justamente associadas ao aumento de importância que os dados ganham no

ambiente de negócios. Conforme pôde ser visto nesta pesquisa, oportunidades e benefícios oriundos do *Big Data* podem ser encontradas nos mais diversos segmentos e indústrias de mercado, tanto no setor público como no privado. Analogamente, oportunidades e benefícios oriundos do *Big Data* também podem ser encontrados nos mais diversos departamentos de quaisquer organizações. Por conseguinte, mais importante do que enunciar as potencialidades da tecnologia é evidenciar os benefícios gerais e os principais ganhos a nível gerencial e de gestão com os quais a tecnologia pode contribuir para a organização, descritos nesse trabalho e recuperados na tabela a seguir.

Autores	Benefícios			
	Aumento da eficiência operacional e melhor desempenho financeiro	Melhora na tomada de decisão	Inovação incremental em produtos e serviços já existentes, e aumento da lealdade dos clientes	Inovação disruptiva de produtos, serviços e modelos de negócios
Bloem <i>et. al.</i> (2012)			x	
Brown, Chui e Manyika (2011)		x	x	
Davenport (2014)	x	x	x	x
Gallant (2011)		x	x	
Leeflang, Verhoeff <i>et. al.</i> (2014)	x		x	
McAfee e Brynjolfsson (2012)	x		x	x
McKinsey Global Inst. (2011)	x	x		x
Minelli <i>et. al.</i> (2013)	x	x		
Novo e Neves (2013)	x		x	x
Ohlhorst (2013)			x	
Silva & Campos (2013)	x			x
Tankard (2012)	x	x		x
Ularu, Puican <i>et. al.</i> (2012)			x	x
Yan (2013)	x			

Para ter acesso a esses benefícios, no entanto, as organizações interessadas devem estar cientes dos fatores relevantes na adoção e uso do *Big Data* no contexto organizacional, o que nos faz retornar ao objetivo geral desta pesquisa, que foi identificar os fatores relevantes na adoção e o uso de *Big Data* no PENSA, órgão da prefeitura municipal do Rio de Janeiro.

Após extensa revisão bibliográfica e análise das entrevistas dos dados coletados, proponho a existência de fatores relevantes relacionados a adoção e uso de *Big Data*, reunidos nas proposições a seguir.

<b>Proposição 1</b>	<b>Quatro fatores organizacionais são relevantes na implementação e uso de <i>Big Data</i> no setor público: o alinhamento à estratégia da organização; o patrocínio da alta gestão; o alinhamento entre os processos organizacionais e o sistema de <i>Big Data</i>; e o envolvimento e retenção de funcionários qualificados em análises de dados.</b>
1.1	O alinhamento estratégico da implementação do <i>Big Data</i> é relevante na adoção e uso do sistema no setor público, pois identifica quais necessidades serão atendidas, quais capacitações devem ser desenvolvidas e estabelece objetivos claros, mensuráveis, realizáveis e verificáveis, norteando e sustentando o compromisso de todos.
1.2	O envolvimento da alta gestão na implementação e uso do <i>Big Data</i> é importante tanto para fins de suporte e patrocínio à nova tecnologia, como para fins de comunicação da estratégia de adoção e auxílio na superação de barreiras à mudança.
1.3	O alinhamento entre os processos organizacionais e o <i>Big Data</i> é um fator relevante na implementação e uso no setor público, pois diminui o impacto da tecnologia no <i>status quo</i> e a resistência à implementação, além de auxiliar no desenvolvimento de uma cultura mais aberta à exploração dos dados e na superação dos desafios oriundos da integração dos silos de dados.
1.4	O envolvimento e retenção de funcionários qualificados em análises de dados é um fator relevante na adoção e uso do <i>Big Data</i> no setor público. A busca de profissionais nas instituições de ensino superior e a formação interior de talentos são soluções para o <i>Talent Gap</i> no médio e longo prazos. No curto prazo, a abordagem de equipe é uma solução eficaz para reunir as competências necessárias para execução de análises de dados.

<b>Proposição 2</b>	<b>Três fatores relacionados à implementação são relevantes para adoção e uso de <i>Big Data</i> no setor público: o uso de estruturas facilitadoras da implementação; o uso de métodos de implementação ágil; e o uso de uma abordagem faseada e evolucionária.</b>
2.1	O uso de pelo menos uma estrutura facilitadora à implementação – comitê gestor, líder/campeão de projetos e grupos de superusuários – é relevante na adoção e uso de <i>Big Data</i> .
2.2	O método de implementação de projetos ágil é relevante para à implementação e uso de <i>Big Data</i> , sobretudo por sua capacidade apresentar resultados rápidos, que aumenta a percepção de valor dos usuários.
2.3	A abordagem faseada é um fator relevante na adoção e uso do <i>Big Data</i> , tendo em vista que ela favorece uma implementação gradual e por meio de protótipos junto aos demais departamentos da organização, contribuindo para a gestão das expectativas e dos desafios, que mudam ao longo do tempo.

<b>Proposição 3</b>	<b>Dos fatores relacionados à percepção de valor do <i>Big Data</i>, destacam-se como fatores relevantes a experimentação, a comunicação, a ética e a privacidade.</b>
3.1	<p>As expectativas dos atores envolvidos variam ao longo da implementação e uso do <i>Big Data</i>. No setor público, a gestão das expectativas dá lugar a experimentação dos diferentes tipos de análise possíveis, na busca de insights, é um fator de geração de valor na adoção e no uso de <i>Big Data</i>.</p>
3.2	<p>A comunicação é um fator relevante na implementação e uso do <i>Big Data</i>. A comunicação deve ocorrer desde a promoção do projeto até a resolução de conflitos internos, pois contribui para a compreensão de todas as capacidades do sistema. A comunicação contínua estimula a troca de informações e deve ocorrer entre todos os stakeholders e deve informar o progresso do projeto.</p>
3.3	<p>O monitoramento e a avaliação de resultados não é um fator de percepção de valor relevante na implementação e uso do <i>Big Data</i> no setor público no Brasil. O uso de mecanismos de implementação pode contribuir para garantir a eficácia da implementação, mas não implica em uma percepção de valor tal que o monitoramento em si seja relevante na implementação, pois os resultados podem ser avaliados qualitativamente.</p>
3.4	<p>A ética é um fator de percepção de valor relevante na implementação e uso de <i>Big Data</i> no setor público no Brasil. Um planejamento prévio deve ser feitos de maneira a impor limites éticos, discutir quais dados podem ser analisados e sob que circunstâncias. Devem ser apontados os responsáveis por decidir sobre questões éticas que não forem previamente analisadas.</p>
3.5	<p>O respeito à Privacidade é um fator de percepção de valor relevante na implementação e uso de <i>Big Data</i> no setor público no Brasil. A garantia de transparência sobre os métodos de coleta e análise de dados é a melhor maneira de lidar com questões sensíveis. A Segurança, por outro lado, não é um fator de percepção de valor relevantes considerando o mesmo contexto.</p>

<b>Proposição 4</b>	<b>Existem fatores tecnológicos e/ou associados à tecnologia que são relevantes na implementação e uso do <i>Big Data</i>. Entre eles se destacam fatores associados à estrutura da equipe de análise de dados em relação à organização, a infraestrutura de sistemas de análise de dados e os desafios relacionados à obtenção dos dados para análise.</b>
4.1	<p>A estrutura do departamento de análises é um fator relevante na implementação e uso do <i>Big Data</i>. A estrutura deve ser constantemente repensada para aumentar a captação de benefícios oriundos da análise de dados, e seu planejamento deve considerar o contexto e a cultura.</p>
4.2	<p>Há evidências de que o planejamento e escolha da infraestrutura de análise de dados são fatores relevantes na adoção e uso do <i>Big Data</i>. O uso de softwares livres e <i>open source</i> e de hardwares capazes de atender às necessidades de análises de dados favorece a percepção de valor.</p>
4.3	<p>A disponibilidade dos dados e de registros históricos, bem como a completude dos dados e uma boa estrutura de metadados são fatores relevantes na implementação e uso do <i>Big Data</i>. A cultura de geração de dados completos é importante para que haja análises de qualidade.</p>

Os fatores relevantes organizacionais são o alinhamento estratégico, o envolvimento e patrocínio da alta gestão, o alinhamento entre os processos organizacionais e o sistema de *Big Data*, e o envolvimento e a retenção de funcionários qualificados. Destes, é interessante destacar que o envolvimento de funcionários qualificados não se refere tão somente à contratação de cientistas de dados, podendo ser ao invés disso a contratação de diversos funcionários multidisciplinares dotados de competências distintas e complementares.

Os fatores relevantes relacionados tão somente à implementação são o uso de pelo menos uma estrutura facilitadora de implementação, o uso de métodos de implementação ágeis e o uso de uma abordagem faseada e evolucionária. Os fatores relevantes relacionados à percepção e valor são a comunicação, ética e privacidade.

Outros fatores como gestão de expectativas, o monitoramento e a avaliação de resultados, e a segurança não se mostraram relevantes na adoção e uso de *Big Data* no caso em questão, diferindo um pouco do que é apontado pela literatura acadêmica. A gestão da expectativa parece ter dado lugar para uma cultura de experimentação, onde as potencialidades da análise de dados são reveladas durante a realização dos projetos de *analytics*.

A ausência destes fatores que pareciam relevantes para a adoção e o uso de *Big Data*, pode ser dar por diversos motivos, como por exemplo o fato de o PENSEA ser um departamento pertencente a uma organização do setor público do Brasil. Embora, devido ao método de pesquisa desenvolvido, não seja possível generalizar os resultados encontrados, pesquisas futuras podem tentar identificar o porquê tais fatores não foram relevantes na adoção do *Big Data*, tendo sido amplamente reconhecidos pela literatura acadêmica como fatores relevantes na adoção de sistemas de BI. Será uma particularidade do contexto analisado no caso ou uma peculiaridade da tecnologia?

Os fatores relevantes relacionados à tecnologia foram à estrutura do departamento de análise de dados, o planejamento e a escolha da infraestrutura tecnológica de análise de dados e os próprios dados em si. Os fatores tecnológicos foram a maior surpresa oriunda da análise das entrevistas, pois não foram fatores identificados na revisão bibliográfica com exceção de Evers (2014), que, no entanto, não os detalhou. Embora possam ter aparecido por questões contextuais ou associadas ao método de pesquisa, trabalhos

futuros podem identificar os principais aspectos relacionados a esses fatores de modo a detalhar melhor a relevância destes fatores.

Com base nos resultados encontrados, as implicações gerenciais que ficam para quaisquer organizações que estudem a possibilidade de aderir ao *Big Data*, sobretudo aquelas que pertencem ao setor público, são que os fatores relevantes na adoção e uso do *Big Data* não são apenas relacionados à implementação em si, mas também a aspectos organizacionais, tecnológicos e de percepção de valor da tecnologia. Os aspectos relacionados à implementação, à percepção de valor e as questões organizacionais permitem que os gestores possam planejar a adoção de maneira a minimizar as chances de insucesso, enquanto que os fatores tecnológicos permitem que os gestores possam planejar a estrutura e a infraestrutura para evitar o surgimento de barreiras e desafios que não possam ser superados, tais como inexistência de dados para analisar ou inexistência de dados completos e de qualidade.

Por fim, vale lembrar que tal qual qualquer outro estudo, essa pesquisa também possui algumas limitações, algumas delas relacionadas ao método de pesquisa, tal como a impossibilidade de generalizar estatisticamente os resultados, a dificuldade para determinar causas e efeitos de eventos reconstruídos, entre outros. No entanto, os resultados aqui dispostos e explicados são válidos respeitando essa janela temporal, tendo em vista que ao longo do tempo, com o desenvolvimento e da maturidade do *Big Data*, os fatores relevantes para sua adoção e uso podem diferir. Ainda assim, eles ficam de aprendizado.

## BIBLIOGRAFIA

ANDRADE, G. Big Data no trânsito: Rio ganha painéis eletrônicos com dados do Waze. **Techtudo**, 2014. Disponível em: <<http://www.techtudo.com.br/noticias/noticia/2014/10/big-data-no-transito-rio-ganha-paineis-eletronicos-com-dados-do-waze.html>>. Acesso em: Junho 2015.

ANSARINEJAD, A. et al. Evaluating the Critical Success Factors in ERP Implementation using Fuzzy AHP Approach. **International Journal of Academic Research**, 3, January 2011.

BARRENCE, A.; GOMES, D. T.; FREITAS, E. E. Dataviva: Big Data em 100 milhões de visualizações abertas sobre a economia brasileira. **Harvard Business Review**, Março 2014.

BARRET, M.; WALSHAM, G. Electronic trading and work transformation in the London insurance market. **Information System Research**, 1999.

BATTY, M. Big data, smart cities and city planning. **Dialogues in Human Geography**, 3, 10 dez. 2013. 274-279. Disponível em: <<http://dhg.sagepub.com/content/3/3/274>>. Acesso em: Março 2015.

BAZZOTTI, C.; GARCIA, E. A Importância do sistema de informação gerencial na Gestão empresarial para tomada de decisões. **Ciências Sociais Aplicadas em Revista**, 2006.

BEATH, C. et al. Finding value in the information explosion. **MIT Sloan Management Review**, 53, 2012. 18.

BEAUDRY, A.; PINSONNEAULT, A. Understanding Users responses to Information Technology: A Coping Model of User Adaption. **Mis Quarterly**, v. 29, n. 3, p. 493-534, Setembro 2005.

BEDER, S. Controversy and closure: Sydney's beaches in crisis. **Social Studies of Science**, 21, May 1991. 223-256.

BERGER, P. L.; LUCKMAN, T. **The Social Construction of Reality**. New York: Anchor Books, 1966.

BIG DATA STARTUPS. Trucking company US Xpress drives efficiency with Big data. **Big data startups**, 2012. Disponível em: <<http://www.bigdata->



startups.com/BigData-startup/trucking-company-xpress-drives-efficiency-big-data/>.

Acesso em: 2 Maio 2015.

BINGI, P.; SHARMA, M. K.; GODLA, J. K. Critical Issues affecting an ERP Implementation.. **IS Management**, 1999.

BIRNEY, E. The making of ENCODE: Lesson for big-data projects. **Nature**, v. 489, p. 49-51, September 2012.

BLOEM, J. et al. **Big Social: Predicting behavior with Big data**. Sogeti. Norway. 2012. (2).

BLOOMBERG. The Current State of Business Analytics: Where do we go from here? **Bloomberg Businessweek Research Services**, 2011. Disponível em: <[http://www.sas.com/resources/asset/busanalyticsstudy\\_wp\\_08232011.pdf](http://www.sas.com/resources/asset/busanalyticsstudy_wp_08232011.pdf)>. Acesso em: 11 Junho 2014.

BLOOMBERG. Research Update—Driving Revenues with Big Data. **Bloomberg Businessweek Research Services**, August 2012.

BLUMENSCHHEIN, F. Estratégias de Compras Governamentais no Brasil: Teoria dos Leilões e "Big data". **FGV Projetos**, v. 21, 2014.

BORKOVICH, D. J.; MORRIS, R. The Social Science of Data Warehousing: Its ever-evolving corporate culture. **Issues in Informations Systems**, XII, 2011. 23-35.

BOURDIEU, J. **Razões práticas: sobre a teoria da ação**. Campinas: Papirus, 2005.

BOYD, D.; CRAWFORD, K. Critical questions for Big Data: provocations for a cultural, technological and scholarly phenomenon. **Information, Communication & Society**, 2012.

BOYD, D.; MARWICK, A. Social privacy in networked publics: teens attitudes, practices, and strategies. **Privacy Law Scholars Conference**, 2 June 2011. Disponível em: <<http://www.danah.org/papers/2011/SocialPrivacyPLSC-Draft.pdf>>. Acesso em: 20 Julho 2014.

BROWN, B.; CHUI, M.; MANYIKA, J. Are you ready for the era of 'big data'? **McKinsey Quarterly**, v. 4, p. 1-12, October 2011.

BUGHIN, J.; LIVINGSTON, J.; MARWAHA, S. Seizing the Potential of 'big data'. **McKinsey Quarterly**, October 2011. Disponível em:

<<http://whispersandshouts.typepad.com/files/using-big-data-to-drive-strategy-and-innovation.pdf>>. Acesso em: June 2014.

BURTON, R. M.; MASTRANGELO, D.; SALVADOR, F. Big Data and Organization Design. **Journal of Organizational Design**, 3, 2014.

CANARY, V. P. **A Tomada de Decisão no contexto do Big data: Estudo de caso único**. Universidade Federal do Rio Grande do Sul. Porto Alegre. 2013.

CEARLEY, D.; CLAUNCH, C. **The Top 10 Strategic Technology Trends**, 2012. Disponível em: <[smama.ch](http://smama.ch)>. Acesso em: 20 Julho 2014.

CHEN, C. L. P.; ZHANG, C.-Y. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. **Information Sciences**, January 2014. Disponível em: <<http://dx.doi.org/10.1016/j.ins.2014.01.015>>.

CHEN, H. Smart Health and Well-being. **IEEE Intelligent Systems**, v. 26, p. 78-79, 2011. ISSN 5.

CHEN, H.; CHIANG, R. H. L.; STOREY, V. C. Business Intelligence and Analytics: From Big data to Big impact. **Mis Quarterly**, v. 36, p. 1-22, December 2012. ISSN 4.

CLIFFE, S. ERP implementation-How to avoid \$100 million write-offs. **Harvard Business School**, 1999.

COPELAND, E. Big Data in the Big Apple, 2014. Disponível em: <<http://capitalcityfoundation.london/bigdatainthebigapplewebversion/>>. Acesso em: July 2015.

CRESWELL, J. Qualitative inquiry and research design: Choosing among five traditions, Thousand Oaks, CA, 1998.

DAVENPORT, ; PATIL, D. J. Data Scientist: the sexiest job of the 21st century. **Harvard Business School**, 2012.

DAVENPORT, T. **Big data at work: dispelling the myths, uncovering the opportunities**. Boston: Harvard Business School Publishing Corporation, 2014.

DAVENPORT, T. H. **Competing on Analytics**. [S.l.]: Harvard Business School, 2006.

DAVENPORT, T.; BARTH, P.; BEAN, R. How 'Big data' is different. **MIT Sloan Management Review**, July 2012.

DEAN, J.; GHEMAWAT, S. MapReduce: Simplified Data Processing on large clusters. **Communications of the ACM**, 2008.

DEMPSEY, J. X.; FLINT, L. M. Commercial Data and National Security. **Washington Law Review**, 1459, 2004. 1474.

DEVLIN, B.; ROGERS, S.; MYERS, J. **Big Data Comes of Age**. EMA and 9sight Consulting Report. [S.l.]. 2012.

DHOLAKIA, R. R.; DHOLAKIA, . **Scholarly Research in Marketing: Trends and Challenges in the Era of the Big Data**. [S.l.]. 2013.

DOMINGUES, B. Seis passos para confiar em Big data. **Blog**, 2013. Disponível em: <<http://www.baguete.com.br/artigos/1268/bruno-domingues//06/08/2013/seis-passos-que-sua-empresa-deve-seguir-para-confiar-em-Big-data>>. Acesso em: maio 2015.

ESS, C. Ethical decision-making and Internet research: Recommendations from the aoir ethics working committee. **Association of Internet Researchers**, 2002. Disponível em: <<http://aoir.org/reports/ethics.pdf>>. Acesso em: 14 Julho 2014.

EVERS, J. M. **Critical Success Factors of Business Intelligence and Big data Analysis: utilizing hidden value from Big data for Business Intelligence**. [S.l.]. 2014.

FANG, ; PATRECIA,. Critical Success Factors in ERP Implementation. **Jönköping International Business School**, 2005. Disponível em: <[diva-portal.org](http://diva-portal.org)>. Acesso em: 14 July 2014.

FAYET, L. 7 Implementation Challenges of Big Data Analytics. **LinkedIn**, 2015. Disponível em: <<https://www.linkedin.com/pulse/7implementationchallengesbigdataanalyticslaurentfayet>>. Acesso em: 05 May 2015.

FINNEY, S.; CORBETT, M. ERP Implementation: A Compilation and Analysis of Critical Success Factors. **Business Process Management Journal**, 2007.

FITCHARD, K. Despite Critics, Cisco Stands by Its Data Deluge. **GigaOM**, 2012. Disponível em: <<http://gigaom.com/2012/02/14/despite-critics-cisco-stands-by-its-data-deluge/>>. Acesso em: May 2013.

FLOYER, D. et al. Big data database Revenue and Market forecast 2012-2017. **Wikibon.org**, 2013. Disponível em: <[http://wikibon.org/wiki/v/Big\\_Data\\_Database\\_Revenue\\_and\\_Market\\_Forecast\\_2012-2017](http://wikibon.org/wiki/v/Big_Data_Database_Revenue_and_Market_Forecast_2012-2017)>. Acesso em: 2 Maio 2015.

FREEMAN, L. 2013 in review: Big data, bigger expectations? **Computer World**, 2013. Disponível em: <<http://blogs.computerworld.com/big-data/23266/2013-review-big-data-cfbdcw>>. Acesso em: July 2014.

GALLANT, TIBCO CEO: How Real-Time Computing Will Change the Landscape. **ComputeWorld**, 2011.

GALLIERS, R. D.; LEIDNER, D. E. **Strategic Information Management: Challenges and Strategies in Managing Information Systems**. 3. ed. Oxford: Taylor & Francis Group, 2003.

GANTZ, J.; REINSEL, D. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. **IDC iView: IDC Analyze the Future**, n. v.2007, p. 1-16, 2012.

GILSINAN, K.; STEPAN, A. From Compstat to Gov 2.0: Big Data in New York City Management. **Case Consortium @ Columbia**, New York, 2014.

GODOY, A. S. Introdução à pesquisa qualitativa e suas possibilidades. **Revista administração empresarial**, v. 35, p. 57-63, 1995. ISSN 2.

GROVES, et al. The 'big data' revolution in healthcare: Accelerating value and innovation. **McKinsey & Company**, 2013. Disponível em: <[www.mckinsey.com](http://www.mckinsey.com)>. Acesso em: 20 Junho 2014.

GUEST, ; BUNCE, ; JOHNSON,. How Many Interviews Are Enough? An Experiment with Data Saturation and Variability Field Methods, 18, 2006. 59-82.

GUPTA, R.; GUPTA, S.; SINGHAL, A. Big Data: Overview. **International Journal of Computer Trends and Technology**, 9, 2014.

HOLLAND, C. P.; LIGHT, B.; GIBSON, N. A Critical Success Factors Model for Enterprise Resource Planning Implementation. **ECIS**, 1999. Disponível em: <<http://www.christopherholland.co.uk/>>. Acesso em: 28 Maio 2014.

IBM. **THE 2011 IBM TECH TRENDS REPORT: Tech Trends of today. Skills for tomorrow.** [S.l.]. 2011.

INFOCHIMPS. How to Do a Big Data Project, 2013. Disponível em: <<http://cdn2.hubspot.net/hub/174427/file-41128102-pdf/docs/bigdataprojecthowto.pdf?t=1386826264000>>. Acesso em: 1 Maio 2015. Site alternativo: <http://www.infochimps.com/resources/how-to-do-a-big-data-project-a-template-for-success/>.

INSTITUTE FOR ADVANCED ANALYTICS. Tools for Managing and Measuring the Value, Winston-Salem, 2014. Disponível em: <<http://analytics.ncsu.edu/sesug/2014/PSA-07.pdf>>. Acesso em: 04 de junho 2015.

KAISLER, S. et al. **Big Data: Issues and Challenges Moving Forward.** 46th Hawaii International Conference on System Sciences. [S.l.]. 2013.

KIM, G.-H.; TRIMI, S.; CHUNG, J.-H. Big Data Applications in the Government Sector. **Communications of the ACM**, v. 57, March 2014. ISSN 3.

LABRINIDIS, ; JAGADISH, H. V. Challenges and Opportunities with Big data. **Proceedings of the VLDB Endowment**, 5, 2012. Disponível em: <[vldb.org](http://vldb.org)>. Acesso em: Junho 2014.

LANEY, D. 3-D Data Management: Controlling Data Volume, Velocity and Variety. **META Group Research Note**, 2001. Disponível em: <<http://goo.gl/Bo3GS>>. Acesso em: junho 2014.

LEEFLANG, P. S. H. et al. Challenges and Solutions for Marketing in a Digital Era. **European Management Journal**, 2014. 1-12. Disponível em: <<http://goo.gl/z1uBzT>>. Acesso em: Maio 2014.

LEININGER, D. Big Data: The Proof of Concept Project. **Developer.com**, 2015. Disponível em: <[www.developer.com/db/big-data-the-proof-of-concept-project.html](http://www.developer.com/db/big-data-the-proof-of-concept-project.html)>. Acesso em: junho 2015.

LI, E. Y. Perceived Importance of Information System Success Factors: a meta analysis of group differences. **Information & Management**, 1997.

LYON, D. Surveillance, Snowden, and Big data: Capacities, Consequences, critique. **Big data & Society**, July-September 2014. 1-13.

LYYTIEN, K.; HIRSCHHEIM, R. Information Systems Failures: A Survey and Classification of the Empirical Literature. **Oxford Surveys in Information Technology**, 1987. 257-309.

MAÇADA, A. C. G.; CANARY, V. P. A Tomada de Decisão no Contexto do Big data: Estudo de caso único. **XXXVIII Enanpad 2014**, Rio de Janeiro, 13-17 Setembro 2014.

MAJCHRZAK, A.; COTTON, J. A Longitudinal Study of Adjustment to Technological Change: From to Computer-Automated Batch Production. **Journal of Occupational Psychology**, v. 61, p. 43-66, 1988. ISSN 1.

MANYIKA, J. et al. **Big data: The next frontier for innovation, competition, and productivity**. McKinsey. [S.l.]. 2011.

MATHEUS, R. Untitled Prezi. **Prezi**, 2013. Disponível em: <[https://prezi.com/vodczw\\_os-\\_o/untitled-prezi/](https://prezi.com/vodczw_os-_o/untitled-prezi/)>. Acesso em: Junho 2015. Apresentação sobre o que o PENSA disponível nos arquivos do Prezi.

MCAFEE, A.; BRYNJOLFSSON, E. Big Data: The Management Revolution. **Harvard Business Review**, p. 59-68, 2012.

MCGUIRE, T.; MANYIKA, J.; CHUI, M. WHY BIG DATA IS THE NEW COMPETITIVE ADVANTAGE. **Ivey Business Journal**, July/August 2012. Disponível em: <<http://iveybusinessjournal.com/publication/why-big-data-is-the-new-competitive-advantage/#.UQhF447C86w>>. Acesso em: April 2015.

MCKINSEY GLOBAL INSTITUTE. Big Data: The next frontier for innovation, competition and productivity. **McKinsey**, 2011. Disponível em: <[www.mckinsey.com/Insights/MGI/Research/Technology\\_and\\_Innovation/Big\\_data\\_The\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation)>. Acesso em: 5 March 2015.

MICHAEL, K.; MILLER, K. W. Big Data: New Opportunities and New Challenges. **IEEE Computer Society**, 2013.

MILLER, C. C. Data Science: the numbers of our lives. **The New York Times**, New York, Abril 2013. Disponível em: <[biostat.jhsph.edu](http://biostat.jhsph.edu)>. Acesso em: Maio 2014.

MINELLI, M.; CHAMBERS, M.; DHIRAJ, A. **Big Data Big Analytics: Emerging Business Intelligence and Analytic trends for today's businesses**. Hoboken: John Wiley and Sons Inc., 2013.

MURRAY, M.; COFFIN, G. A Case Study analysis of factors for success in ERP System implementations. **Proceedings of the Seventh Americas Conference on Information Systems**, Boston, 2001. 1012–1018.

NAH, F. F. H.; LAU, J. L. S.; KUANG, J. Critical Factors for Successful Implementation of Enterprise Systems. **Emerald**, 2001. Disponível em: <<http://www.emerald-library.com/ft>>.

NEIVA, R. Estrangeiro pode prestar concurso público? **Exame.com**, 27 dez. 2011. Disponível em: <<http://exame.abril.com.br/carreira/noticias/estrangeiropodeprestarconcursopublico>>. Acesso em: 2015.

NOVO, R.; NEVES, J. M. S. D. **Inovação na inteligência analítica por meio do Big data**: característica de diferenciação da abordagem tradicional. VIII Workshop de Pós-graduação e Pesquisa do Centro Paula Souza. São Paulo: [s.n.]. 2013. p. 32-44.

OHLHORST, F. **Big Data Analytics**: Turning Big data into Big money. [S.l.]: Wiley, 2013.

OLIVEIRA, D. D. P. R. D. **Sistemas, organização e métodos**: uma abordagem gerencial. 13. ed. São Paulo: Atlas, 2002.

PATTERSON, D. J. Helping Protists to Find Their Place in a Big Data World. **Acta Protozoologica**, p. 115-128, 2014.

PIPINO, L. L.; YANG, W. L.; WANG, R. Y. Data Quality Assessment. **Communications of the ACM**, 45, April 2002.

POON, P. P.; WAGNER, C. Critical Success Factors revisited: Success and Failure Cases of Information Systems for Senior Executives. **Decision Support Systems**, 2001.

PORTO, F. **Big Data in Astronomy: The LIneA-DEXL case**. EMC/NCE/UFRJ. Rio de Janeiro. 2013. Palestra apresentada no 1o. EMC summer School on Big data.

RAGHUPATHI, W.; RAGHUPATHI, V. Big data analytics in healthcare: promise and potential. **Health Information Science and Systems**, v. 2, 2014. Disponível em: <<http://www.hissjournal.com/content/2/1/3>>.

RIBEIRO, C. J. S. **Big data**: os novos desafios para o profissional da informação. Informação & Tecnologia (ITEC). João Pessoa/Marília: [s.n.]. 2014. p. 96-105.

- ROBB, D. 9 Tips on Using Big Data Analytics to Improve Security. **Enterprise APPS Today**, 2014. Disponível em: <<http://www.enterpriseappstoday.com/datamanagement/9-tips-on-using-big-data-analytics-to-improve-security.html>>. Acesso em: 28 May 2015.
- ROBERTS, H. J.; BARRAR, P. R. N. MRPII implementation: key factors for success. **Computer Integrated Manufacturing Systems**, 1992.
- ROBINSON, A. G.; DILTS, D. M. OR & ERP: a match for the new millennium? **OR/MS Today**, 26, 1999.
- RODRIGUES, E. **A interação entre Sistemas de informação e o Trabalho no Setor bancário no Brasil: uma análise estruturalista**. [S.l.]: [s.n.], 2008.
- RUSSOM, P. Big Data Analytics. **TDWI Best Practices Report**, 2011. Disponível em: <[tdwi.org](http://tdwi.org)>. Acesso em: Maio 2014.
- SANTOS, I. H. R. **BigData Research and Development at Petrobras**. EMC/NCE/UFRJ. Rio de Janeiro. 2014.
- SCHMARZO, B. **Big data: understanding how data powers Big Business**. Indianapolis: John Wiley & Sons, Inc., 2013.
- SHANKS, G. et al. Differences in critical success factors in ERP systems implementation in Australia and China: a cultural analysis. **Proceedings of the 8th European Conference on Information Systems**, 2000.
- SHUEH, J. Big Data Could Bring Governments Big Benefits. **Governing**, 2014. Disponível em: <[www.governing.com/templates/gov\\_print\\_article?id=251538951](http://www.governing.com/templates/gov_print_article?id=251538951)>. Acesso em: 21 Maio 2015.
- SHVACHKO, K. et al. **The Hadoop distributed file system**. 26th IEEE Symposium on Mass Storage Systems and Technologies. [S.l.]: [s.n.]. 2010.
- SICULAR,. Gartners' Big Data definition consists of three parts, not to be confused with Three "V's". **Gartner Inc.**, 2013. Disponível em: <<http://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/>>. Acesso em: Junho 2014.



SILVA, I. M. D.; CAMPOS, F. C. D. **New Perspectives using Big data: a study of bibliometric 2000-2012**. 11th International Conference on Information Systems and Technology Management - CONTECSI. São Paulo: [s.n.]. 2014.

SIMON, P. **Too Big too Ignore**. Hoboken: John Wiley & sons Inc., 2013.

SKOK, W.; LEGGE, M. Evaluating enterprise resource planning systems using an interpretive approach. **Proceedings of the 2001 ACM SIGCPR conference on Computer Personnel Research**, San Diego, April 2001. 189-197.

SOMERS, T. M.; NELSON, ; RAGOWSKY,. Enterprise resource planning (ERP) for the next milenium: development of an integrative framework and implications for research. **Proceedings of the Americas Conference on Information Systems (AMCIS)**, 2001. 998-1004.

SOUZA, L. Big Data vira política pública no Rio. **Baguete**, 2014. Disponível em: <<http://www.baguete.com.br/noticias/24/09/2014/big-data-vira-politica-publica-no-rio>>. Acesso em: Junho 2015.

STAMFORD, C. Gartner Says Big Data Will Drive \$28 Billion of IT Spending in 2012. **Gartner**, 2012. Disponível em: <<http://www.gartner.com/newsroom/id/2200815>>. Acesso em: 15 Maio 2015.

STEFANOUC, C. J. Supply chain management (SCM) and organizational key factors for successful implementation of enterprise resource planning (ERP) systems. **Proceedings of the Americas Conference on Information Systems (AMCIS)**, 1999. 800.

SUMNER, M. Critical success factors in enterprise wide information management systems projects. **Proceedings of the Americas Conference on Information Systems (AMCIS)**, 1999. 232-234.

TANKARD, C. Big data security. **Netwrok security**, p. 5-8, July 2012.

TAURION,. Você realmente sabe o que é Big data? **IBM**, 12 Agosto 2012. Disponível em: <[https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/voce\\_realment\\_e\\_sabe\\_o\\_que\\_e\\_big\\_data?lang=en](https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/voce_realment_e_sabe_o_que_e_big_data?lang=en)>. Acesso em: 2 Maio 2015.

TENE, O.; POLONETSKY, J. Big Data for All: Privacy and User Control in the age of Analytics. **Northwestern Journal of Technology and Intellectual Property**, 11, Abril 2013.

THE U.S. OFFICE OF PERSONNEL MANAGEMENT. Open Government version 2.0., May 2012. Disponível em: <[http://www.opm.gov/about-us/open-government/reference-materials/governmentplan\\_v20.pdf](http://www.opm.gov/about-us/open-government/reference-materials/governmentplan_v20.pdf)>.

THE WHITE HOUSE. Big data is a big deal, 29 March 2012. Disponível em: <<http://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal>>.

THIRY-CHERQUES, H. R. **Métodos estruturalistas**: pesquisas em ciências de gestão. São Paulo: Atlas, 2008a.

TOLE, A. A. Big Data Challenges. **Database Systems Journal**, IV, 2013. Acesso em: May 2015.

TYRE, M. J.; ORLIKOWSKI, W. J. Windows of Opportunity: Temporal Patterns of Technological Adaptation in Organizations. **Organizatio Science**, v. 5, p. 98-118, 1994. ISSN 1.

TYRE, M. J.; ORLIKOWSKI, W. J. The Episodic Process of Learning by Using. **International Journal of Technology Management**, v. 11, p. 790-798, 1996. ISSN 7/8.

ULARU, E. G. et al. Perspectives on Big Data and Big Data Analytics. **Database System Journal**, 3, 2012.

UMBLE, J. E.; UMBLE, M. M. Avoiding ERP implementation failure. **Industrial Management**, 44, 2002. 25-35.

VAN RIJMENAM, M. 4 Benefits For The Public Sector When Governments Start Using Big Data. **Datafloq**, 2014. Disponível em: <<https://datafloq.com/read/4benefitspublicsectorgovernmentsstartbigdat/171>>. Acesso em: 28 Maio 2015.

VEJA RIO. Rio tem "Tropa de Elite" para Big Data. **Veja Rio**, 2014. Disponível em: <<http://vejario.abril.com.br/materia/cidade/rio-tem-tropa-elite-big-data/>>. Acesso em: Junho 2015.

WACTLAR, H.; PAVEL, M.; BARKIS, W. Can Computer Science Save Healthcare?, v. 6, p. 79-83, 2011. ISSN 5.

WEBER, K.; OTTO, B.; OSTERLE, H. One Size Does Not Fit All - A Contingency Approach to Data Governance. **ACM Journal of Data and Information Quality**, 1, June 2009. 2-27.

XEXÉO, G. Big data: computação para uma sociedade conectada e digitalizada. **Ciência Hoje**, v. 306, p. 19-23, Agosto 2013. Disponível em: <[http://cienciahoje.uol.com.br/revista-ch/2013/306/pdf\\_aberto/bigdata306.pdf](http://cienciahoje.uol.com.br/revista-ch/2013/306/pdf_aberto/bigdata306.pdf)>. Acesso em: 1 Maio 2014.

YAN, J. **Big data, Bigger opportunities**. 2012 President Management Council Inter-agency Rotation Program. [S.l.]: [s.n.]. 2013.

YEO, K. T. Critical Failure Factors in Information System Projects. **International Journal of Project Management**, 2002.

YEOH, W.; KORONIUS,. Critical success factors for business intelligence systems. **Journal of Computer Information Systems**, 2010. 23-32.

YIN, R. K. **Case Study Research: Design and Methods**. [S.l.]: Sage Publications Inc., 1989.

ZANNI, P. P.; MORAES, G. H. S. M. D.; MARIOTTO, F. L. Para que servem os Estudos de Caso Único? **XXXV Encontro da ANPAD**, Rio de Janeiro, Setembro 2011.

ZIKOPOULOS, P. et al. **Harness the Power of Big data: The IBM Big data Platform**. McGraw Hill Professional. [S.l.]. 2012.

## ANEXO I – ROTEIRO DE PESQUISA

### ROTEIRO DE PESQUISA

O presente roteiro de pesquisa é feito para a realização de uma entrevista com integrantes e ex-integrantes da equipe do P3NS4. As perguntas em negrito correspondem ao tema e são feitas aos integrantes e então lhes é dado tempo para responder. Dependendo da resposta dos entrevistados, são então realizadas as perguntas em itálico para que a resposta seja mais completa dentro dos fins da pesquisa.

#### **Tema: Início do projeto**

---

**Você pode me contar um pouco sobre como foi à implantação de *Big Data* na prefeitura com a criação do P3NS4?**

1. *Qual era o objetivo inicial do projeto?*
2. *Como você se envolveu com o P3NS4? Qual era a sua participação?*
3. *Em sua opinião, havia alinhamento da estratégia da prefeitura com as atividades do P3NS4?*
4. *Qual a importância do “relacionamento”, tanto externo (com outras secretarias e com fornecedores de tecnologia) quanto interno (dentro da equipe) para o sucesso das atividades?*
5. *Como funcionava a equipe? Cada um possuía a sua função?*
6. *Qual era o grau de dedicação da equipe?*
7. *Havia envolvimento da alta gestão da prefeitura no P3NS4? De que maneiras? Quanto importante era esse envolvimento?*
8. *Alguém na equipe já tinha experiência em Big Data?*
9. *Como era tratada a identificação de novas oportunidades?*

#### **Tema: Evolução do projeto**

---

**Como estava o Big Data/P3NS4 quando você se desligou da equipe?**

1. *Havia ocorrido desdobramentos com relação à proposição inicial do P3NS4?*
2. *Havia ocorrido mudanças de escopo do projeto inicial? Se sim, por quê?*
3. *Que departamentos e secretarias faziam uso dos serviços do P3NS4 até aquele momento?*
4. *Como eram selecionados os projetos a serem executados? Internamente ou por demanda externa?*
5. *Havia alguma gestão de expectativas sobre as potencialidades do P3NS4 e do Big Data?*
6. *Havia algum líder ou campeão de projeto? Isto é, alguém responsável por “vender” o P3NS4 à aqueles que podiam se beneficiar dele.*
7. *Com relação à equipe inicial, alguém havia sido contratado posteriormente? Se sim, como e porque esse novo funcionário foi selecionado?*
8. *Você diria que o P3NS4 foi desenvolvido em fases? Se sim, as fases foram planejadas previamente? Qual foi a importância disso?*

### **Tema: Mudança Organizacional**

---

#### **Como foi feita a gestão da mudança organizacional para a implementação do P3NS4?**

1. *Você diria que houve influência do P3NS4 nos processos organizacionais da prefeitura e das secretarias que trabalharam com o P3NS4?*
2. *Houve alguma alteração nas relações interdepartamentais/estrutura organizacional?*
3. *No desenvolvimento das atividades e estudos, houve algum conflito com alguma outra área? Se sim, conflito de que tipo? Como foi desenvolvida a solução?*
4. *Houve problema com relação à propriedade da informação? Isto é, ocorreram os chamados “silos de informação”?*
5. *Havia demanda por novas competências de conhecimento?*
6. *Como era realizada a conscientização e o treinamento dos funcionários com relação às técnicas de análise de dados? Havia transferência de conhecimento?*
7. *Havia superusuários do P3NS4?*
8. *Na realização das atividades, havia oferta de recompensas financeiras e não-financeiras de acordo com o alcance de metas?*

### **Tema: Impactos do Big Data**

---

#### **Qual o impacto que o P3NS4/Big Data teve na prefeitura?**

1. *Como eram os processos de tomada de decisões na prefeitura antes da implementação do P3NS4? E depois?*
2. *Os insights gerados pelo Big Data e pela equipe eram validados de alguma forma?*
3. *Houve dificuldades técnicas e gerenciais em algum dos projetos? Se sim, quais? (exemplos: dificuldades de relacionamento gestor/funcionário ou falta de conhecimento)*
4. *Como eram avaliados os custos, riscos e benefícios dos projetos?*
5. *Do P3NS4 como um todo, como podem ser avaliados os custos, riscos e benefícios?*

### **Tema: Ética e privacidade**

---

#### **Quais as preocupações de natureza ética existem em projetos como o P3NS4?**

1. *Havia preocupações como o anonimato dos indivíduos?*
2. *Como eram determinados quais dados podiam ser usados nas análises?*
3. *Quem era responsável por avaliar a ética no uso dos dados?*
4. *Havia compartilhamento de dados com terceiros?*
5. *Havia algum armazenamento da proveniência de dados? Ex: metadado em geral.*

### **Tema: Questões técnicas gerais**

---

1. *Havia alinhamento entre as ações técnicas e o alcance dos objetivos de negócios?*
2. *Qual estilo de planejamento de projetos era utilizado? Em cascata, progressivo ou ágil? Qual o impacto dessa opção?*
3. *Qual foi o critério de escolha da arquitetura tecnológica? Ele era adequado?*
4. *Quem foi consultado na escolha da arquitetura tecnológica?*
5. *O uso de tecnologias open source influenciou na escolha?*

6. *Quão flexível era a arquitetura para atender a necessidades previstas de Big Data?*
7. *Quem pagava pelos custos da infraestrutura? TI, a área-fim ou a própria equipe?*
8. *Tiveram algum problema com o tempo de processamento dos modelos?*
9. *Tiveram algum problema com confiabilidade da infraestrutura?*
10. *Houve alguma integração do P3NS4 com os sistemas legados das secretarias?*
11. *Houve alguma preocupação em melhorar a qualidade (ex.: informação não tão confiável)?*
12. *Houve necessidade de análise em tempo real ou quase real? Se sim, foi utilizado algum pré-processamento para acelerar a análise?*
13. *Houve algum problema com o volume ou falta de estrutura dos dados?*
14. *Havia requisitos na visualização de dados*

**Você gostaria de fazer algum comentário extra?**

**Muito obrigado pela sua contribuição à pesquisa.**

## ANEXO II – LISTA DE DEFINIÇÕES DE *BIG DATA*

Lista de definição de *Big Data* por autores. As dimensões V1, V2, V3, V4, V5, O e T são, respectivamente, volume, velocidade, variedade, valor, veracidade, outras dimensões e tecnologias de apoio.

Autores	Definição adotada	V1	V2	V3	V4	V5	O	T
Ammu & Irfanuddin (2013)	"Big Data é uma frase popular usada para descrever um enorme volume de dados estruturados e não estruturados, que é tão grande que é difícil processar com técnicas tradicionais de banco de dados e software tradicionais. As características que em geral distinguem Big Data são os 3 V's: volume, variedade e maiores taxas de velocidade"	x	x	x				x
Arnold (2012)	"Pode ser definido como quantidades massivas de conteúdo armazenado (estruturado ou não) que pode ser facilmente analisado em tempo real (em uma quantidade de tempo razoável para alcançar uma resposta útil)."	x	x	x				
Batty (2013)	"Big data é 'qualquer conjunto de dado que não pode ser inserido em uma planilha de Excel'"	x	x	x				
BEATH et al. (2012)	"Desafios de armazenar, proteger e acessar quantidades massivas de dados"	x						
Begoli & Horey (2012)	"Refere-se a prática de coletar e processar conjuntos muito grandes de dados e uso de sistemas associados e algoritmos para a análise desses conjuntos massivos de dados"	x					x	
Bhatia & Vaswani (2013)	Há inúmeras maneiras de mensurar Big data - que podem ser baseadas em volume, variedade, velocidade e valor.	x	x	x	x			
Bonometti (2012)	"A totalidade de fontes de dados estruturados e não estruturados, incluso mídias sociais"	x	x	x				
BUGHIN, CHUI, MANYIKA (2010)	Inundação de dados "como resultado de maior acesso de dados do consumidor através de fontes públicas, de proprietário e pagas, bem como novas informações advindas de comunidades web e de novas implantações de ativos inteligentes"	x		x				
Canary (2013)	Volume, velocidade, variedade, valor, veracidade	x	x	x	x	x	x	
Chang, Kauffman et al. (2013)	Volume, velocidade, variedade, valor	x	x	x	x			

Autores	Definição adotada	V1	V2	V3	V4	V5	O	T
Chen & Zhang (2014)	<i>Big Data é uma coleção de conjuntos de dados muito grandes com uma grande diversidade de tipos de modo que torna-se difícil processo usando abordagens de processamento de dados mais modernas ou plataformas de processamento de dados tradicionais.</i>	x	x	x				x
Chen, Chiang & Storey (2012)	<i>"Descreve os conjuntos de dados e técnicas analíticas em aplicações que são tão grandes (de terabytes para exabytes) e complexas (de sensor a dados de mídias sociais) que eles requerem únicas e avançadas tecnologias de armazenamento de dados, administração, análise e visualização."</i>	x	x	x				x
Cumbley & Church (2013)	<i>Grande parte do debate sobre Big Data tem sido impulsionada pelo tamanho (...) Impressionante como são, grande quantidade de dados eletrônicos é de pouca utilidade em si. Uma parte importante do movimento de Big Data são as novas tecnologias que vem sendo usadas para extrair informações significativas (...) volumes de dados atuais não estão sendo impulsionadas por dados estruturados tradicionais, mas por uma explosão de dados não estruturados ou semi-estruturados.</i>	x		x				x
Davenport (2012)	<i>Dados demasiadamente volumosos ou muito desestruturados para serem gerenciados e analisados através de meios tradicionais</i>	x		x				x
Deloitte (2013)	<i>Volume, Velocidade, variedade, veracidade, viabilidade, valor</i>	x	x	x	x	x	x	
Emerald Group (2013)	<i>"Identifica[...] Big Data [...] como os 3 V's: volume, variedade e velocidade"</i>	x	x	x				



Autores	Definição adotada	V1	V2	V3	V4	V5	O	T
Evers (2014)	<i>"Big Data, de um ponto de vista organizacional, são dados de múltiplas fontes e bases de dados, incluindo estruturado, bem como dados não estruturados. (...) Do ponto de vista tecnológico, Big data engloba grandes volumes, variedade e velocidade de dados que não podem ser corretamente processados e analisadas por meio de técnicas convencionais, e, assim, inclui técnicas avançadas necessárias para processar esses dados, tais como o processamento paralelo."</i>	x	x	x				x
Fernandes, O'Connor & Weaver (2015)	<i>"Big Data pode ser definida por suas características básicas, por vezes referida como os 3 V's: volume, velocidade e variedade"</i>	x	x	x				
Freitas Junior & Maçada (2014)	<i>Observa-se que foram apresentadas citações de autores mais recentes, que vêm estudando o tema nos últimos três anos, assim, com base no quadro acima, pode-se notar uma similaridade nos conceitos, que podem ser sintetizados como sendo um grande volume de dados estruturados ou não, de fontes diversas, que devem ser gerenciados e analisados de forma peculiar.</i>	x		x				x
French (2012)	<i>"A disponibilidade de vastas quantidades de dados de todos os tipos e o aumento de velocidade e poder de máquinas para analisar esses dados"</i>	x	x	x				
Gantz & Reinsel (2012)	<i>"'Tecnologias de Big Data' descreve uma nova geração de tecnologias e arquiteturas, concebidas para extrair economicamente valor de volumes muito grandes de uma ampla variedade de dados, permitindo a alta velocidade de captura, descoberta, e / ou análise."</i>	x	x	x	x			x
Gartner Group (2012)	<i>"Big Data, em geral, é definido como ativos de alto volume, velocidade e variedade de informação que exigem custo-benefício, de formas inovadoras de processamento de informações para maior visibilidade e tomada de decisão."</i>	x	x	x			x	

Autores	Definição adotada	V1	V2	V3	V4	V5	O	T
Gobble (2013)	"Quando são tão grandes [os dados] que os sistemas tradicionais não conseguem lidar."	x					x	
Gordon-Murdane (2012)	"Não é apenas aumento na quantidade e tipos de dados, é também melhores ferramentas para armazenar, agregar, combinar, analisar, e extrair novas idéias."	x					x	
Griffin (2012)	"É importante notar que big data é muito mais do que mídias sociais. É dados estruturados e não estruturados residindo em bases de dados em múltiplas regiões geográficas. É texto em formulários Web e PDFs, e é email e todas as formas de outros documentos."			x				
Grunes & Stcke (2015)	Big data é frequentemente caracterizada por quatro "Vs": volume, velocidade, variedade e valor.	x	x	x	x			
HBR (2013)	"Big Data" é um termo usado para descrever o novo volume, variedade e velocidade de dados que agora enfrenta empresas	x	x	x				
IBM (2014)	"Volume, velocidade, variedade e veracidade"	x	x	x		x		
Jacobs (2009)	"Big data deve ser definida em qualquer ponto no tempo, como 'dados cujo tamanho nos obriga a olhar para além dos métodos experimentados e verdadeiros que são predominantes naquela época'. Hoje em dia, isso pode significar que os dados é muito grande para ser colocado em um banco de dados relacional"	x						
Kaisler, Armour at al. (2013)	"Características de Big Data:" variedade, volume, velocidade, valor, complexidade	x	x	x	x		x	
Katal, Wazid & Goudar (2013)	"Big data pode ser definida com as seguintes propriedades associadas a ele:" variedade, volume, velocidade, variabilidade, complexidade, valor	x	x	x	x		x	x
Kraska (2013)	"É quando o aplicativo normal da tecnologia atual não permite aos usuários obter no tempo correto, com custo eficaz e com respostas de qualidade a perguntas orientadas por dados."	x	x	x				

Autores	Definição adotada	V1	V2	V3	V4	V5	O	T
Kumar & Vidhyalakshmi (2012)	"Volume, variedade e velocidade são os três principais fatores de Big Data"	x	x	x				
Kwon, Lee & Shin (2014)	"Caracterizado em volume, variedade, velocidade e valor, 'big data' é considerado por profissionais da indústria como o próximo 'blue ocean' de oportunidades de negócios. Definimos grandes tecnologias de análise de dados (por exemplo, banco de dados e ferramentas de mineração de dados) e técnicas (por exemplo, métodos analíticos) que uma empresa pode empregar para analisar dados em grande escala, dados complexos para várias aplicações de modo a aumentar o desempenho da empresa em várias dimensões"	x	x	x	x			x
Leeflang, Verhoef et al. (2014)	"Por big data queremos dizer conjuntos de dados tão grande e complexo que se torna difícil para processar usando ferramentas de gerenciamento de banco de dados tradicionais ou aplicações de processamento de dado"	x	x	x				x
Lima Junior (2012)	"Big data" refere-se ao conjunto de dados (dataset) cujo tamanho está além da habilidade de ferramentas típicas de banco de dados em capturar, gerenciar e analisar."	x						x
Lohr (2012)	"Refere-se à inundação crescente de dados digitais de muitas fontes, incluindo a internet, sensores biológicos e industriais, videos, email e redes sociais de comunicação."	x		x				
Lopez (2012)	"Existem muitas definições de big data , mas a maioria dos especialistas concordam em três características fundamentais: volume, velocidade e variedade. Outro aspecto chave, frequentemente negligenciado, é custo."	x	x	x				
Lyon (2014)	"...huge volume, consisting of terabytes or petabytes of data; high velocity, being created in or near real time; extensive variety, both structured and unstructured;..."	x	x	x				
Maçada & Canary (2013)	Apresentado em tabela	x	x	x	x	x		

Autores	Definição adotada	V1	V2	V3	V4	V5	O	T
Madden (2012)	"Dados que são em grandes quantidades, muito rápidos e muito difíceis para as ferramentas atuais de processamento"	x	x	x			x	
Mahrt & Scharkow (2013)	"Denotam um maior conjunto de dados ao longo do tempo, conjuntos de dados estes que são grandes demais para serem manipulados por infraestruturas de armazenamento e processamento regulares"	x						x
MANYIKA et al. (2011)	"Conjunto de dados cujo tamanho está além da habilidade de típicas ferramentas de software de base de dados para capturar, armazenar, administrar e analisar"							
Marshall (2012)	"Muito grande para ser observado de perto"	x						
Michael & Miller (2013)	"Big Data reúne não apenas grandes quantidades de dados, mas também vários tipos de dados que anteriormente nunca teria sido consideradas em conjunto. Estes fluxos de dados exigem cada vez maior velocidade de processamento, mas devem ser armazenados economicamente"	x	x	x				x
Minelli et al. (2013)	"A indústria tem uma definição evoluindo em torno de Big Data que está definido por três dimensões: 1. Volume 2. Variedade 3. Velocity"	x	x	x				
More, Chaudhary et al. (2013)	"'Big Data' refere-se a conjuntos de dados, cujo tamanho é além da capacidade de captura, armazenagem, gerenciamento e análise do típico software de banco de dados"	x						x
NewVantagePartners (2013)	"Big Data é um termo usado para descrever conjuntos de dados tão grandes, tão complexos ou que requeiram tratamento rápido (...) que se tornam difícil ou impossível de trabalhar com o uso de gestão de banco de dados padrão ou ferramentas analíticas. Manipulando conjuntos de dados como estes muitas vezes exigem software maciçamente paralelo em execução em dezenas, centenas ou até milhares de servidores"	x	x	x				x

Autores	Definição adotada	V1	V2	V3	V4	V5	O	T
Novo & Neves (2013)	"Comparando os exemplos aqui listados é possível identificar as três dimensões do big data (volume, velocidade e variedade) presentes"	x	x	x				
Pospiech & Feldens (2012)	"Um desafio de uma relação desfavorável entre dados disponíveis e tecnologias de informação ou conceitos atuais."	x					x	
Preimesberger (2011)	"Um crescente número de conjunto de dados empresariais de tamanhos colossais e toda a tecnologia necessária para criar, armazenar, cruzar, analisar, arquivar e recuperar esses dados empresariais."	x	x	x			x	
PwC (2014)	A recente onda de informação eletrônica produzida em maior volume por um número crescente de fontes (ou seja, não apenas os dados coletados por uma organização particular no curso de negócios normal).	x		x				
Raghupathi & Raghupathi (2014)	"Por definição, Big data em saúde refere-se a dados de saúde eletrônicos tão grandes e complexos que é difícil (ou impossível) de gerenciar com software e / ou hardware tradicional; nem podem ser facilmente gerenciadas com ferramentas e métodos de gestão de dados tradicionais ou comuns"	x	x	x				x
Ribeiro (2014)	A abordagem de Big Data está apoiada em quatro outros fatores de sustentação, conhecidos como os 4 Vs do Big Data: Volume, Variedade, Velocidade e Veracidade (...) Uso de tecnologias específicas, tais como processamento de rotinas em paralelo e ferramentas para otimização como Hadoop e MapReduce, HDFS, além de abordagens de MachineLearning e Analytics.	x	x	x	x			x
Rogers (2011)	"Conjuntos de dados que não podem mais ser facilmente administrados ou analisados com ferramentas, métodos ou infraestruturas tradicionais ou comuns de administração de dados."			x			x	

Autores	Definição adotada	V1	V2	V3	V4	V5	O	T
Silva & Campos (2014)	<i>As definições existentes na literatura para o Big Data convergem para os seguintes fatos, a utilização de diferentes fontes, tipos de dados e características que se refere ao volume, variedade e velocidade</i>	x	x	x				
Singh & Singh (2012)	<i>"Conjunto de dados que continuam a crescer tanto que torna difícil de administrá-los usando conceitos e ferramentas existentes de administração de base de dados."</i>	x					x	
Tabuena (2012)	<i>"Big data não é apenas sobre tamanho. Big data é realmente sobre ferramentas de dados do tipo analytics."</i>	x		x				
Tankard et al (2012)	<i>"Refere-se à quantidade de informações cada vez maiores que as organizações estão armazenando, processando e analisando, devido ao crescente número das fontes de informações em uso."</i>	x		x				
Taurion (2012)	<i>O que é Big Data? Outro dia escrevi um post com uma fórmula simples para conceitualizá-lo. Big Data = volume + variedade + velocidade. Hoje adiciono mais dois "V"s: veracidade e valor.</i>	x	x	x	x	x		
TechAmerica Foundation (2012)	<i>"Big Data é um termo que descreve grandes volumes de alta velocidade, complexo e variáveis de dados que exigem técnicas e tecnologias avançadas para permitir a captura, armazenamento, distribuição, gestão e análise da informação"</i>	x	x	x			x	
Tole (2013)	<i>"Os 3 V's de Laney [volume, velocidade e variedade] [...] representam elementos-chave que são considerados vitais sobre as características dos sistemas de Big Data. (...) Depois dos 3 V's de Laney, "mais dois" V's de [valor e veracidade] foram adicionados como aspectos fundamentais dos sistemas de Big Data".</i>	x	x	x	x	x		

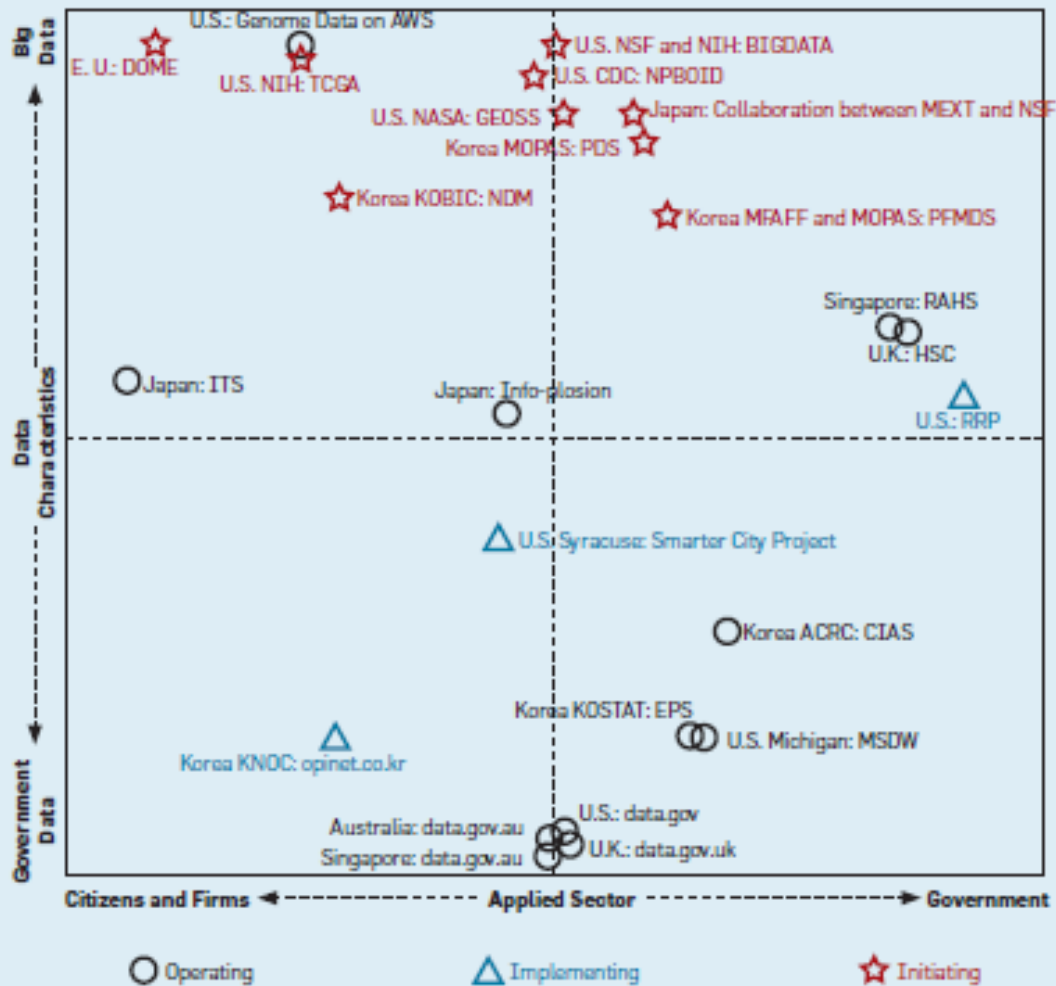
Autores	Definição adotada	V1	V2	V3	V4	V5	O	T
Ularu, Puican et al. (2012)	<i>Há uma série de definições sobre Big Data em circulação no mundo, mas consideramos que o mais importante é aquela que cada líder dá aos dados de sua empresa. A maneira que Big Data é definida tem implicação na estratégia de uma empresa. Cada líder tem que definir o conceito, a fim de trazer vantagem competitiva para a empresa.</i>	x	x	x	x	x	x	x
Villars, Olofson & Eastwood (2011)	<i>(...) é possível que alguns funcionários brilhante em sua organização lidem efetivamente com a variedade muitos dados, volume, e os problemas de velocidade. Por definição, a taxa de crescimento Big Data excede as capacidades de infra-estrutura de TI tradicional e representa grande parte de computação e gerenciamento de dados de problemas para os clientes.</i>	x	x	x				x
Xexeo (2013)	<i>Há consenso de que três dessas características, as iniciadas pelos três 'Vs', são as principais: volume, velocidade e variedade (...) A partir desses três 'Vs', diversos autores propõem ainda outros conceitos, como veracidade, variabilidade ou valor.</i>	x	x	x	x	x	x	
Yan (2013)	<i>Um esclarecimento a ser feito nas discussões sobre o conceito de Big data é que o termo pode se referir tanto a grandes e/ou diversas bases de dados, ou para tecnologias de lidar com esses tipos de conjuntos de dados.</i>	x		x				x

## ANEXO III – EVOLUÇÃO CONCEITUAL E TEMPORAL DE SISTEMAS DE ANÁLISE DE DADOS

<b>BI&amp;A 1.0</b> <i>Conteúdo estruturado, baseados em tecnologia DBMS</i>	• RDBMS e data warehousing	<b>Decision Support</b>	<b>Executive Support</b>	<b>OLAP</b> <b>Business Intelligence</b>					
	• Extract, Transform, Load								
	• Análise estatística descritiva								
	• Dashboards e scorecards								
	• Mineração de dados								
	• OLAP								
	• Visualização Iterativa								
	• <i>Ad hoc query</i> baseada em BI								
	• Análise preditiva								
<b>BI&amp;A 2.0</b> <i>Conteúdo desestruturado, baseado na WEB</i>	• Extração e recuperação de informação								
	• Simulação e Otimização								
	• Análise espaço-temporal								
	• Resposta de linguagem natural à questionamentos								
	• Análise da Web e Inteligência da Web								
	• Análise de texto e conteúdo								
	• Análise de mídia e social								
	• Mineração de opinião								
	• Decisão em Tempo Real								
<b>BI&amp;A 3.0</b> <i>Conteúdo baseado em sensores e mobile</i> <i>Baseada em tecnologia móvel</i>	• Análise baseada na localização								
	• Análise centrada no indivíduo								
	• Análise de relevância do contexto								
	• Visualização móvel e HCI								
	1970-1975	1975-1980	1980-1985	1985-1990	1990-1995	1995-2000	2000-2005	2005-2010	2010-...



# ANEXO IV – PROJETOS DE *BIG DATA* AO REDOR DO MUNDO



**Japan** Collaboration of Ministry of Education, Culture, Sports, Science, and Technology and National Science Foundation

**Japan** Intelligent Traffic System

**Korea** Anti-Corruption and Civil Rights Commission of Korea: Complaints Information Analysis Center

**Korea** Statistics Korea: Employment Position Statistics

**Korea** Ministry for Food, Agriculture, Forestry, and Fisheries and Ministry of Public Administration and Security: Preventing Foot and Mouth Disease Syndrome System

**Korea** Ministry of Public Administration and Security: Preventing Disasters System

**Singapore** Risk Assessment and Horizon Scanning

**U.K.** Horizon Scanning Center

**U.S.** Centers for Disease Control and Prevention: Networked Phylogenomics for Bacteria and Outbreak ID

**U.S.** Genome Data on Amazon Web Services

**U.S.** Michigan: Michigan Statewide Data Warehouse

**U.S.** National Aeronautics and Space Administration: Global Earth Observation System of Systems

**U.S.** National Science Foundation and National Institutes of Health: BIGDATA

**U.S.** Return in Review

# ANEXO V – DECRETO DE CRIAÇÃO DO MODA (MODELO DE NOVA YORK)



THE CITY OF NEW YORK  
OFFICE OF THE MAYOR  
NEW YORK, N. Y. 10007

EXECUTIVE ORDER No. 306

MAYOR'S OFFICE OF DATA ANALYTICS

April 17, 2013

WHEREAS, the City of New York has become a national model for collecting data to measure government operations; and

WHEREAS, City agencies routinely collect various types of data on the buildings, streets, infrastructure, businesses and other entities within the City, including but not limited to tax records, building permits, crime-related data, noise and other 311 complaints; and

WHEREAS, much of the data collected by the City is stored within each collecting agency, impeding the ability to aggregate, analyze and synthesize it to better allocate public resources; and

WHEREAS, the establishment of an Office of Data Analytics and a centralized data sharing and analysis capacity will enable the City to aggregate and analyze data from across City agencies and other sources to more effectively address crime, public safety, and quality-of-life issues by prioritizing risk more strategically, delivering services more efficiently, and enforcing laws more effectively; and

WHEREAS, under Local Law 11 of 2012 (the "Open Data Law"), the City of New York must make relevant city data publicly available online, giving priority to data sets that can be used to increase agency accountability and responsiveness, improve public knowledge of the agency and its operations, further the mission of City agencies, create economic opportunity, or respond to a need or demand identified by public consultation; and

WHEREAS, the establishment of a centralized data sharing and analysis capacity will assist in the achievement of the objectives of the Open Data Law;

NOW, THEREFORE, by the power vested in me as Mayor of the City of New York, it is hereby ordered that:

Section 1. There is established in the Office of the Mayor the Mayor's Office of Data Analytics (the "Office").

§ 2. The Office will be headed by a Chief Analytics Officer. The Chief Analytics Officer shall report to the Mayor's Chief Advisor for Policy and Strategic Planning and shall consult regularly with the Deputy Mayor for Operations, the Deputy Mayor for Economic Development, and the Commissioner of the Department of Information Technology and Telecommunications (DoITT).

§ 3. The Office shall develop and work with agencies to implement data-driven solutions to City service delivery issues. The Office's responsibilities shall include but not be limited to the following:

- a. Collaborative, Data-Driven Solutions. The Office shall work with City agencies to identify how data held by those agencies can be analyzed and combined with other agencies' data to best fulfill their respective missions, shall develop strategies based on such data, and shall assist agencies in implementing those strategies.
- b. Citywide Data Platform. The Office shall develop and implement a citywide data platform that aggregates and updates data from City and other governmental agencies and other sources and that connects and synthesizes data regarding a single address, business or individual that was previously isolated within individual agencies. In collaboration with DoITT, the Office shall:
  1. Work with city agencies to ensure that they have proper technology to provide and retrieve data from the citywide data platform;
  2. Train agency staff to use the citywide data platform; and
  3. Develop practices for performing ongoing and new data analytics with each agency.
- c. Oversight of Data Projects. The Office shall, as appropriate, oversee agency data projects to ensure that agencies use best-practices and such projects are appropriately prioritized by need, impact and feasibility.
- d. Data liaison duties. The Office shall serve as the designated point of contact for outside partners contributing to or using City data.
- e. Implementation of the Open Data Law. The Office shall work with DoITT to ensure compliance by City agencies with the Open Data Law. The Chief Analytics Officer will also serve as the City's Chief Open Platform Officer (COPO) as defined in DoITT's Open Data Policy and Technical Standards Manual.

§ 4. The Deputy Mayor for Operations shall convene and chair an Analytics Steering Committee consisting of representatives of the Office of Data Analytics, the Commissioner of DoITT, and any other Mayor's Office representative designated by the Deputy Mayor for

Operations. The Steering Committee shall first meet within 30 days of the effective date of this Order and shall develop a citywide analytics strategy.

§ 5. City agencies shall cooperate with the Chief Analytics Officer to ensure appropriate implementation of this Order. Such cooperation shall include but not be limited to the sharing of relevant data in a timely fashion, and agencies shall facilitate and encourage real-time data exchanges whenever possible.

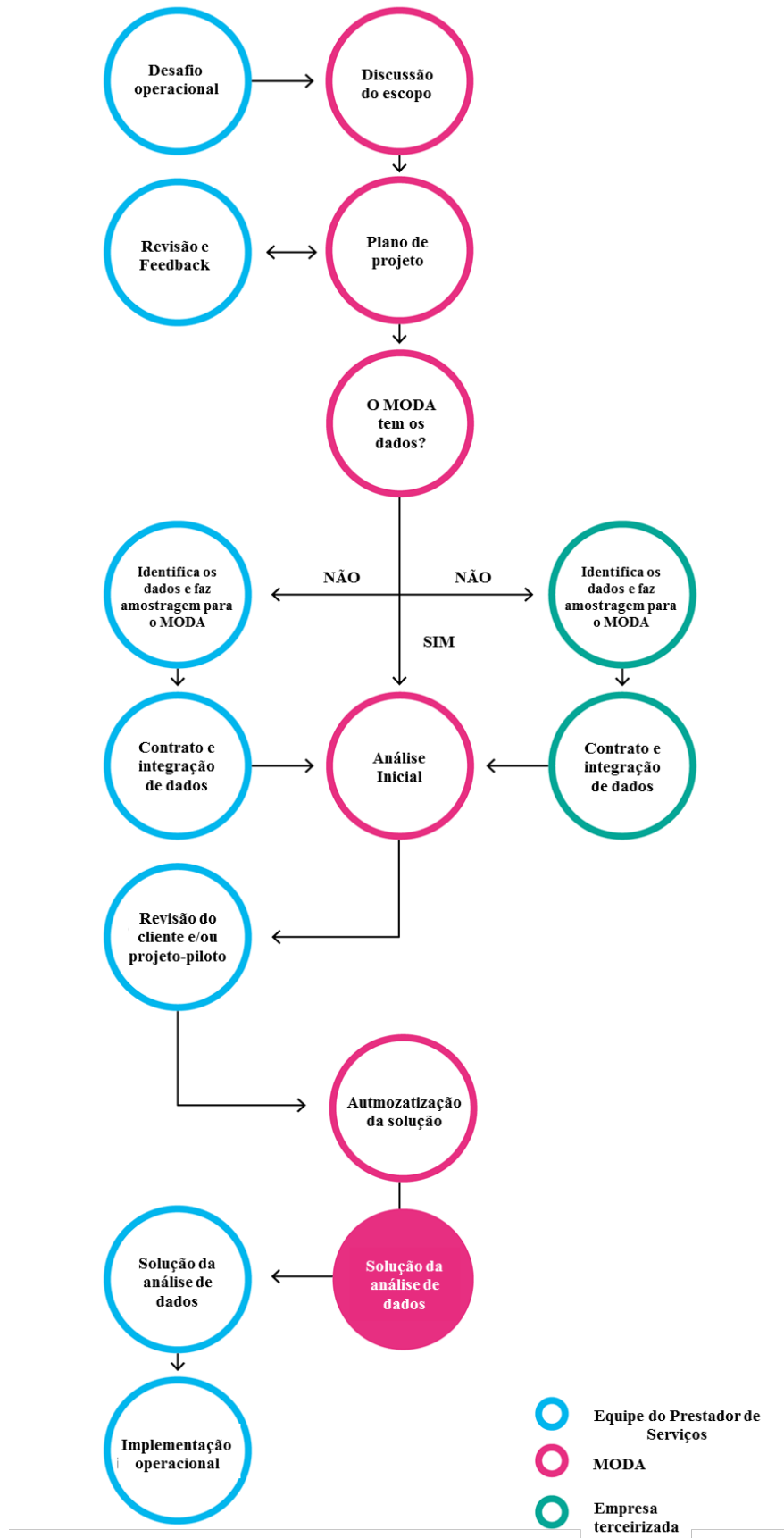
§ 6. Strategies, processes, and solutions developed by the Mayor's Office of Data Analytics, including the Citywide Data Platform, shall be aligned with the Citywide Information Technology Strategy, as well as the security requirements and enterprise architecture standards as set forth by DoITT.

§ 7. This Order shall take effect immediately.



Michael R. Bloomberg  
Mayor

## ANEXO VI - MODELO DE PROCESSOS DO MODA



## ANEXO VII – DECRETO DE CRIAÇÃO DO PENSA

### ATOS DO PREFEITO

DECRETO Nº 37215 DE 3 DE JUNHO DE 2013

Cria a área de Big Data – “PENSA – SALA DE IDEIAS”, no âmbito da Prefeitura da Cidade do Rio de Janeiro.

O PREFEITO DA CIDADE DO RIO DE JANEIRO, no uso de suas atribuições legais,

CONSIDERANDO o novo Modelo de Gestão de Alto Desempenho da Prefeitura do Rio de Janeiro;

CONSIDERANDO a grande gama de serviços prestados pela Prefeitura nos seus diferentes órgãos;

CONSIDERANDO o grande volume de dados gerados, muitos deles não estruturados, tratados pela Prefeitura compartimentalizados por diferentes órgãos;

CONSIDERANDO a complexidade dos problemas e análises a serem resolvidos;

CONSIDERANDO a responsabilidade da Prefeitura de buscar sempre uma prestação mais eficiente e eficaz de serviços públicos ao cidadão;

CONSIDERANDO a transversalidade das políticas públicas e necessidade de buscar sinergias, correlações e alavancas entre informações providas pelos diversos órgãos;

DECRETA:

Art. 1.º Fica criada, na estrutura organizacional da Secretaria Municipal da Casa Civil – CASA CIVIL, a área de BIG DATA: PENSA - SALA DE IDEIAS,

#### CAPÍTULO I Do Objetivo

Art 2º O PENSA - SALA DE IDEIAS tem como finalidade pesquisar, analisar, avaliar correlações e definir ações de impacto a partir do cruzamento dos diferentes bancos de dados disponíveis dentro e fora Prefeitura com o objetivo de aprimorar a prestação de serviços ao cidadão.

Parágrafo único. O PENSA - SALA DE IDEIAS deverá trabalhar em permanente parceria com o Instituto Pereira Passos – IPP e sua respectiva Diretoria de Informações da Cidade e a Empresa Municipal de Informática – IPLANRIO.

#### CAPÍTULO II Das Atribuições

Art 3º Compete a equipe PENSA - SALA DE IDEIAS:

- I - Entender os serviços prestados pela Prefeitura bem como seus desafios;
- II - Trabalhar com a ampla base de dados da Prefeitura, estruturados ou não, oriundos dos diversos órgãos municipais;
- III - Dar suporte de análise de dados aos times dos diferentes órgãos na busca de novas soluções;
- IV – Apontar possíveis soluções aos gestores da prefeitura para problemas;



V - Definir e implementar planos de ação, estruturais e tecnológicos, para ser reconhecido como área de vanguarda em BIG DATA;

VI - Inovar nas soluções propostas;

VII - Criar convênios com universidades, instituições e empresas quando se fizer necessário e que tragam conhecimento e inovação para os processos em questão;

VIII - Formar profissionais da área de BIG DATA.

## **CAPÍTULO II** **Da Estrutura**

Art 4º O PENSE - SALA DE IDEIAS funcionará fisicamente em uma sala especial no Centro de Operações da Prefeitura - COR

Art 5º A equipe de cientistas de dados deverá ser ocupada preferencialmente por profissionais oriundos de áreas técnicas em sua formação com Mestrado e/ou Doutorado.

Art. 6º Este Decreto entra em vigor na data da sua publicação.

Rio de Janeiro, 3 de junho de 2013; 449º ano da fundação da Cidade.

EDUARDO PAES

### **DECRETO Nº 37216 DE 3 DE JUNHO DE 2013**

Modifica o Decreto nº 36.961 de 25 de março de 2013.

O PREFEITO DA CIDADE DO RIO DE JANEIRO, no uso de suas atribuições legais, e tendo em vista o disposto nos artigos 5º, alínea "I", e o 6º do Decreto-Lei n.º 3.365, de 21 de junho de 1941, com a redação que lhe deu a Lei Federal n.º 9.785, de 29 de janeiro de 1999,

DECRETA:

Art. 1º Fica revogado o artigo 4º do Decreto nº 36.961 de 25 de março de 2013.

Art. 2º Este Decreto entra em vigor na data de sua publicação.

Rio de Janeiro, 3 de junho de 2013; 449º ano da fundação da Cidade.

EDUARDO PAES

## ANEXO VIII – DECRETO DE ACESSO AOS DADOS PARA O PENSA

### ATOS DO PREFEITO

#### DECRETO Nº 37540 DE 13 DE AGOSTO DE 2013

Revoga o Decreto nº 37.490, de 07 de agosto de 2013.

O PREFEITO DA CIDADE DO RIO DE JANEIRO, no uso de suas atribuições que lhe são conferidas pela legislação em vigor e,

DECRETA:

Art. 1º Fica revogado o Decreto nº 37.490, de 07 de agosto de 2013, com efeito repristinatório ao Decreto nº 36.968 de 02 de abril de 2013.

Art. 2º Este Decreto entra em vigor na data de sua publicação.  
Rio de Janeiro, 13 de agosto de 2013; 449º ano da fundação da Cidade.  
EDUARDO PAES

#### DECRETO Nº 37541 DE 13 DE AGOSTO DE 2013

Dispõe sobre o acesso aos dados da Administração Direta e Indireta da Cidade do Rio de Janeiro, pela equipe do Big Data - "PENSA - SALA DE IDEIAS".

O PREFEITO DO RIO DE JANEIRO, no uso de suas atribuições que lhe são atribuídas pela legislação em vigor, e

CONSIDERANDO a criação da área de Big Data - "PENSA - SALA DE IDEIAS" pelo Decreto Nº 37.215, de 03 de junho de 2013;

CONSIDERANDO ser necessário para a produção de ações de impacto que a área de Big Data - "PENSA - SALA DE IDEIAS" tenha acesso aos dados produzidos pela Administração Direta e Indireta da Cidade do Rio de Janeiro;

CONSIDERANDO que compete à área de Big Data - "PENSA - SALA DE IDEIAS" trabalhar com a ampla base de dados da Prefeitura, estruturados ou não, oriundos dos diversos órgãos municipais, conforme descrito no inciso II do artigo 3.º do Decreto n.º 37.215/13.

DECRETA:

Art. 1.º Todos os órgãos da Administração Direta e Indireta da Cidade do Rio de Janeiro deverão franquear, de forma célere e não burocrática, o acesso a seus bancos de dados quando solicitado pela equipe do Big Data - "PENSA - SALA DE IDEIAS".

Parágrafo único. Em caso de dúvidas quanto a limites e restrições na disponibilização dos bancos de dados, o órgão solicitado deverá submeter, no prazo de até 5 (cinco) dias, o pedido à Secretaria Municipal da Casa Civil - CVL para decisão, fundamentando as razões para a não liberação dos dados.

Art. 2.º Os órgãos da Administração Direta e Indireta do Município do Rio de Janeiro deverão articular esforços conjuntamente, contando com apoio do Instituto Municipal de Urbanismo Pereira Passos – IPP e sua respectiva Diretoria de Informações da Cidade e a Empresa Municipal de Informática – IPLANRIO para o efetivo cumprimento do disposto neste Decreto.

Art. 3.º Este Decreto entra em vigor na data de sua publicação.  
Rio de Janeiro, 13 de agosto de 2013; 449º ano da fundação da Cidade.  
EDUARDO PAES