



CLEIDE LUCIANE ANTONIUTTI

USOS DO *BIG DATA* EM CAMPANHAS ELEITORAIS

Tese de Doutorado
Agosto de 2015



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO- UFRJ
ESCOLA DE COMUNICAÇÃO – ECO
INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA - IBICT
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO – PPGCI

CLEIDE LUCIANE ANTONIUTTI

USOS DO *BIG DATA* EM CAMPANHAS ELEITORAIS

Rio de Janeiro

2015

CLEIDE LUCIANE ANTONIUTTI

USOS DO *BIG DATA* EM CAMPANHAS ELEITORAIS

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Informação, convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro/Escola de Comunicação, como requisito parcial à obtenção do título de Doutor em Ciência da Informação.

Orientadora: Sarita Albagli

Rio de Janeiro

2015

A635

Antoniutti, Cleide Luciane

Usos do Big Data em campanhas eleitorais / Cleide Luciane
Antoniutti. 2015.

270 f.: il.

Orientador: Prof. Dr. Sarita Albagli.

Tese (doutorado) - Programa de Pós-Graduação em Ciência da
Informação, Instituto Brasileiro de Informação em Ciência e
Tecnologia Universidade Federal do Rio de Janeiro, Escola de
Comunicação, Rio de Janeiro, 2015.

1. Big Data – Aspectos sociais. 2. Campanha eleitoral. 3. Mineração
de dados. I. Albagli, Sarita. II. Universidade Federal do Rio de Janeiro.
Escola de Comunicação. III. Instituto Brasileiro de Informação em
Ciência e Tecnologia. Programa de Pós-Graduação em Ciência da
Informação.

CDD: 005.7

CLEIDE LUCIANE ANTONIUTTI

USOS DO *BIG DATA* EM CAMPANHAS ELEITORAIS

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Informação, convênio entre o Instituto Brasileiro de Informação em Ciência e Tecnologia e a Universidade Federal do Rio de Janeiro/Escola de Comunicação, como requisito parcial à obtenção do título de Doutor em Ciência da Informação.

Aprovado em: 07 de Agosto de 2015.

Profa Dra Sarita Albagli (Orientadora)
PPGCI/IBICT- ECO/UFRJ

Prof Dr Marco André Feldman Schneider
PPGCI/IBICT-ECO/UFRJ

Prof Dr Marcos do Couto Bezerra Cavalcanti
PPGCI/IBICT-ECO/UFRJ
COPPE-UFRJ

Profa Dra Luciana Fernandes Veiga
ECP-UNIRIO

Prof Dr Paulo Eduardo Lins da Silva Cajazeira
IISAC/UFCA

DEDICATÓRIA

Dedico esta tese às pessoas mais importantes da minha vida e que, mesmo nos momentos das minhas ausências ou falta de tempo, nunca deixaram de me amar sem limites.

A minha amada mãe Salette, a quem devo todo o meu incentivo para os estudos.

Ao Carlos César, meu fiel e amado companheiro de todos os momentos nos últimos 6 anos.

Aos meus filhos Anna Luísa e Carlos Eduardo, que são meus presentes dados por Deus.

AGRADECIMENTOS

Quatro anos se passaram e nessa caminhada muitas coisas aconteceram. Passei por momentos de muitas alegrias e conquistas, mas também de muitas angústias e tristezas, principalmente com doenças e a partida inesperada de pessoas da família. E nesse caminho o apoio de muitas pessoas foi fundamental para o incentivo e conclusão desta tese. A lista é grande, mas não posso deixar de agradecer.

A Deus, que tornou possível todas as coisas em minha vida. Por ter me dado saúde e coragem para que eu chegasse até aqui. Agradeço ao Senhor tudo o que sou e tudo o que tenho.

A minha família, que sempre representou a base de tudo e que é a que mais sofreu com as minhas ausências. Sem a paciência e colaboração deles, talvez eu não conseguisse finalizar este trabalho.

A minha mãe Salette Angeli Antoniutti, por me mostrar que estudar é o melhor caminho para conquistar as coisas na vida. Pelas suas orações diárias e pelas muitas lágrimas derramadas por mim nesses 43 anos. Ela é o meu melhor exemplo de mulher, mãe e profissional. Mesmo sozinha e sem nunca ter frequentado o ensino superior conseguiu me guiar até aqui.

Ao meu marido Carlos César Gomes São Braz, que o destino me presenteou e que nesses últimos seis anos têm me mostrado o que é o verdadeiro sentido das palavras casamento e amor. Ele mudou o meu destino e é o maior responsável pelo meu Doutorado em Ciência da Informação nesta instituição. Obrigada por ser esse marido e pai maravilhoso. Por sua dedicação e carinho nas horas de desânimo, cansaço e irritações. E por ser o meu suporte técnico nos momentos de intensa produção científica.

A minha filha Anna Luísa, que se acostumou nos últimos meses com as frases: *“depois que a mamãe terminar o doutorado... depois que a mamãe terminar a tese... depois que a mamãe defender a tese... depois... depois...”*, soube esperar com paciência pelos passeios de fim de semana, pelas idas à praia e ao shopping ou por minutos de atenção e brincadeiras.

Ao meu filho Carlos Eduardo, que veio para colorir de azul a minha vida e que nasceu durante o doutoramento e testou todos os meus limites e capacidades. Desde bebê seguia junto para as aulas do curso e palestras. Por ver tanto a mamãe diante do computador, hoje aos três anos de idade já é um *expert* da tecnologia e um consumidor voraz da informação digital. Ele

desliza seus dedinhos pela tela do *smartphone* à procura dos vídeos do Doki e do Pocoyo (*Discover Kids*) com a maior intimidade tecnológica do momento.

Ao João Pedro, meu enteado, que me ajudou na finalização e nos últimos retoques da tese.

A minha orientadora e professora Dra Sarita Albagli, pela sua confiança e por acreditar na minha pesquisa. Com muita seriedade e competência, conseguiu me guiar pelos melhores caminhos da pesquisa científica para alcançar os objetivos propostos na investigação.

Aos professores do Programa de Pós-Graduação em Ciência da Informação do IBICT – UFRJ, que contribuíram para aprimorar meus conhecimentos na área de Ciência da Informação e me abriram os olhos para muitos caminhos da minha pesquisa. Alguns em especial como a professora Maria Nélide Gonzales de Gomes, que com sua doçura e sua bagagem teórica não mediu esforços para enriquecer a minha temática. À professora Lena Vania Pinheiro, no auge dos seus mais de 70 anos, que, com seu jeito apaixonado de ver as coisas e a vida, sempre sorridente, abriu meus olhos para a Ciência da Informação.

Aos meus colegas de turma pela convivência, troca de experiências profissionais e acadêmicas e mesmos as angústias compartilhadas nos últimos anos. Em especial à amiga Jaqueline Barradas e à Marcia Cavalcanti pelas trocas de ideias, pelas horas dedicadas em ler parte da minha tese. As suas sugestões e contribuições foram fundamentais.

A amiga, professora Dra Claudia Irene Quadros, professora do curso de comunicação social da UFPR (Universidade Federal do Paraná), que sempre me incentivou e me abriu os olhos para a informação, comunicação e tecnologia. Ela plantou a sementinha do tema *big data* numa conversa informal pelo Facebook, me incentivando para a pesquisa.

Ao colega e amigo Márcio Gonçalves pelos incentivos à temática e desabafos da vida acadêmica.

Aos meus amigos que sempre torceram pelo meu sucesso e por esse doutorado.

Aos funcionários do IBICT por todo apoio técnico e burocrático, e em especial a Janete Franco Desidério, que sempre prontamente e carinhosamente atendeu e respondeu as minhas solicitações sobre documentações e outras questões relativas ao curso.

À Sabrina Furtado, que há mais de três anos tem se mostrado uma pessoa leal e amiga, que para nós é mais do que uma simples secretária do lar, é a mãe substituta dos meus filhos nos momentos de ausência. Pelo sua paciência, carinho e dedicação.

À Patrícia Moreira, minha ex-aluna de graduação e orientanda, que por muitas vezes largou tudo para me auxiliar degravando entrevistas e me auxiliando com meus filhos.

Aos entrevistados e aos profissionais que cederam minutos dos seus tempos para falar sobre o tema. Em especial ao jornalista e professor Walter Lima Júnior, ao cientista de dados Maurício Moura, ao especialista em tecnologia da informação César Taurion, ao professor e especialista em marketing político Carlos Manhanelli e ao publicitário Renato Pereira.

Aos profissionais da E.Life (SP) e da Ideia Inteligência (RJ), que abriram suas portas para me mostrar como funcionava o *big data* em suas empresas.

Agradeço em especial aos profissionais que trabalharam no comitê de campanha eleitoral para governo do Rio de Janeiro em 2014, que me acolheram entendendo a importância da minha pesquisa.

Ao ex-deputado federal Basílio Villani, pelos anos de aprendizado como sua assessora de imprensa e marketing, e com quem aprendi a entender e também a tomar gosto por política e campanhas eleitorais.

À banca avaliadora pelas contribuições à pesquisa e pelas futuras sugestões. Espero ter atendido às orientações e sugestões da qualificação no desenvolvimento final desta tese.

À CAPES por ter me financiado com a bolsa de doutoramento durante os quatro anos. Com esse incentivo financeiro, consegui adquirir livros, participar de eventos e mesmo ajudar no sustento da minha casa.

Enfim, ao ensino público, porque desde que me conheço por “gente” sempre ocupei os bancos escolares públicos. Espero ser sempre um exemplo de dedicação para aqueles que almejam chegar aqui.

Você pode chamar de Ciência das Redes, dinâmica humana, computação social ou big data. Não importa o nome que você dê, existe uma quantidade imensa de informações coletadas sobre nós e temos a obrigação de dar um sentido a tudo isso.

Albert-László Barabasi – cientista

ANTONIUTTI, Cleide Luciane. Usos do *big data* em campanhas eleitorais. Orientadora: Sarita Albagli; 2015. 270f. Tese (Doutorado em Ciência da Informação) – Universidade Federal do Rio de Janeiro/Escola de Comunicação, Rio de Janeiro, 2015. Instituto Brasileiro de Informação em Ciência e Tecnologia. Programa de Pós-Graduação em Ciência da Informação.

RESUMO

Esta tese tem como foco os usos que campanhas eleitorais contemporâneas têm feito do *big data*, a partir do fenômeno da datificação da informação e da adoção de novas práticas para conquista de eleitores daí derivadas. O trabalho inicia-se discutindo a informação e a tecnologia a partir da Sociedade Informacional (Castells). Caracteriza em seguida o fenômeno da datificação a partir da disponibilidade e dos usos das mídias sociais digitais e dos dispositivos móveis. Conceitua e caracteriza o *big data* a partir de diferentes dimensões teóricas e práticas, apresentando os impactos do *big data* na vigilância, monitoramento e uso de informações pessoais de ambientes de mídias sociais digitais. Discute as características das campanhas eleitorais atuais e os conceitos de marketing eleitoral. Apresenta campanhas eleitorais centradas em dados, ilustrando com o exemplo da campanha eleitoral americana de Barack Obama de 2012. Na pesquisa empírica, investiga como empresas de dados brasileiras trabalham com o *big data* em campanhas eleitorais. Busca diferentes visões de especialistas sobre a temática. Observa-se diretamente uma campanha eleitoral brasileira no pleito de 2014, que foca suas estratégias no uso do *big data*. Considera-se o *big data* como um processo pouco usado, mas com grandes expectativas, contribuindo para as ações direcionadas e microsegmentadas de comunicação entre candidato e eleitores.

Palavras-chave: Big data; Campanhas Eleitorais; Datificação; Sociedade Informacional; Informação.

ANTONIUTTI, Cleide Luciane. Usos do *big data* em campanhas eleitorais. Orientadora: Sarita Albagli; 2015. 270f. Tese (Doutorado em Ciência da Informação) – Universidade Federal do Rio de Janeiro/Escola de Comunicação, Rio de Janeiro, 2015. Instituto Brasileiro de Informação em Ciência e Tecnologia. Programa de Pós-Graduação em Ciência da Informação.

ABSTRACT

This thesis focuses on the uses that contemporary election campaigns have made of big data, from datificação of information phenomenon and adoption of new practices to achievement of voters derived therefrom. The work begins discussing the information and technology from the Informational Society (Castells). Features then the datificação phenomenon from the availability and use of digital social media and mobile devices. Conceptualizes and features the big data from different theoretical and practical dimensions, presenting the big data impacts on surveillance, monitoring and use of personal information of digital social media environments. Discusses the features of the current election campaigns and the concepts of electoral marketing. Presents election campaigns centered on data, illustrating the example of the American election campaign of Barack Obama 2012. In empirical research investigates how data Brazilian companies work with big data in election campaigns. Search different views of experts on the subject. It is observed directly a Brazilian election campaign in the 2014 election that focuses its strategies on the use of big data. It is considered the big data as a little-used process, but with great expectations, contributing to the targeted and microsegmentadas actions of communication between candidates and voters.

Keywords: Big data; Electoral campaigns; Datificacion; Informational society; Information.

LISTA DE FIGURAS

Figura 1-	Contagem de bytes.....	35
Figura 2-	Volume de Informação digital.....	35
Figura 3-	Pirâmide do Conhecimento.....	40
Figura 4-	Pirâmide de Fluxos e Estoques.....	41
Figura 5 -	Hierarquia Reversa do Conhecimento.....	45
Figura 6-	Bio do Twitter.....	54
Figura 7 -	Três dimensões do big data.....	69
Figura 8 -	Projeto de Bentham.....	89
Figura 9-	Eleitores conectados às redes sociais em 2014.....	138
Figura 10-	Redes sociais mais populares.....	139
Figura 11 -	Mapa das eleições americanas em 2012.....	150
Figura 12-	Dados da Campanha Obama.....	153
Figura 13 -	Anúncio no Orlando Sentinel da Flórida.....	161
Figura 14 -	Página Oficial de Twitter de Obama.....	164
Figura 15 -	Fases da pesquisa de campo.....	168
Figura 16 -	Tipos de pesquisas.....	175
Figura 17 -	Como trabalha a empresa.....	176
Figura 18-	Exemplo de mapa geográfico dos eleitores.....	208
Figura 19 -	Relatórios do Sistema Big Data.....	209
Figura 20 –	Cadastro das ações de rua.....	210
Figura 21 –	Página de Cadastro.....	211
Figura 22 -	Exemplo de cadastro de eleitores.....	212
Figura 23 –	Fórmula de probabilidade.....	213
Figura 24 -	Exemplo de uma base de dados.....	217
Figura 25-	Moderação do Big data.....	217
Figura 26 -	Funil da participação do eleitor nas redes sociais.....	220
Figura 27 -	Alcance do programa de debate eleições.....	221
Figura 28 -	Desempenho dos candidatos em cada bloco do debate.....	221
Figura 29 -	Os influenciadores da campanha.....	222
Figura 30–	Nuvem de palavras.....	223
Figura 31-	Opinião nas redes.....	224

LISTA DE QUADROS

Quadro 1-	Censo populacional brasileiro de 1872.....	50
Quadro 2-	Censos demográficos do Brasil.....	51
Quadro 3-	Mudanças do <i>big data</i>	74
Quadro 4 -	Origem dos dados.....	76
Quadro 5 -	Ferramentas de Monitoramento.....	97
Quadro 6 -	Dados coletados pelo Facebook.....	102
Quadro 7 -	Dados coletados pelo Google.....	102
Quadro 8 -	Dados coletados pelo Instagram.....	103
Quadro 9 -	Dados coletados pelo Waze.....	104
Quadro 10 –	A diferença entre as fases da campanha eleitoral.....	117
Quadro 11-	Estruturação da campanha com dados eleitorais.....	131
Quadro 12 -	Atividades das empresas de dados.....	179
Quadro 13 –	Especialistas entrevistados na pesquisa.....	182
Quadro 14 -	Profissionais entrevistados no Comitê Eleitoral.....	202
Quadro 15 -	Cenário Eleitoral para governo do Rio de Janeiro em 2014.....	203
Quadro 16 –	Temas abordados nas redes.....	225
Quadro 17 –	Comparativo Ocorrências X Alcance.....	225
Quadro 18 -	Os debates no primeiro turno.....	226
Quadro 19 –	Ocorrências X Candidatos.....	226
Quadro 20 –	Comparativo – Alcance X Candidatos.....	227

LISTA DE IMAGENS

Imagem 1 - Painéis de Comunicação Digital.....	84
Imagem 2 - O centro de operações do <i>big data</i> Obama.....	157
Imagem 3 - Estrutura física do Comitê de Campanha.....	205

LISTA DE TABELAS

Tabela 1 -	Definições de <i>big data</i>	68
------------	-------------------------------------	----

LISTA DE SIGLAS

ANP – Agência Nacional do Petróleo
CEO – Chief Executive Officer
CERN – Conseil Européen pour la Recherche Nucléaire
CI – Ciência da Informação
CIA – Central Intelligence Agency
CNPQ - Conselho Nacional de Pesquisa
COPPE - Coordenação de Programas de Pós-Graduação em Engenharia
COR – Centro Operações Rio
CRIE – Centro de Referencial em Inteligência Empresarial
DGE – Diretoria Geral de Estatística
EMC- Empresa dos Estados Unidos de Software y hardware
EUA – Estados Unidos da América
FAPERJ – Fundação de Amparo a Pesquisa do Rio de Janeiro
FBI – Federal Bureau of Investigation
FGV – Fundação Getúlio Vargas
GPS – Global Positioning System
IBGE – Instituto Brasileiro de Geografia e Estatística
IBM – Internation Business Machines
IDC – International Data Corporation
MBI - Master in Business Administration
MGD – Machine – Generated Data
MIT – Massachusetts Institute of Techonology
NASA – National Aeronautics and Space Administration
NSA – Agência Nacional de Segurança dos Estados Unidos
OFA – Obama for America
PMDB – Partido do Movimento Democrático Brasileiro
PRISM – Programa de Espionagem dos Estados Unidos
RFID – Radio-Frequency IDentification
SIME – Sistema de Informação em Marketing Eleitoral
TCCCOG – Tecnologia, Comunicação e Ciência Cognitiva
TI – Tecnologia da Informação
TIC – Tecnologia da Informação e Comunicação
UFPR – Universidade Federal do Paraná
UFRJ – Universidade Federal do Rio de Janeiro
UGC – User Generated Content

SUMÁRIO

1	INTRODUÇÃO	20
1.1	A ESTRUTURA DA TESE	24
	PARTE I – O ENQUADRAMENTO TEÓRICO E TEMÁTICO	26
2	A SOCIEDADE INFORMACIONAL E A DATIFICAÇÃO DA INFORMAÇÃO	27
2.1	O CONTEXTO DA SOCIEDADE INFORMACIONAL	27
2.2	O EXCESSO DE INFORMAÇÃO	29
2.2.1	O acúmulo de dados na era digital	32
2.3	A INFORMAÇÃO COMO CONCEITO INTERDISCIPLINAR E MODERNO.....	37
2.3.1	As relações entre informação, dado e conhecimento	39
2.4	O INÍCIO DA DATIFICAÇÃO	46
2.4.1	O conceito de datificação	52
3	BIG DATA: DIFERENTES DIMENSÕES CONCEITUAIS E PRÁTICAS	57
3.1	BIG DATA NAS CIÊNCIAS SOCIAIS.....	57
3.2	BIG DATA E SUAS PARTICULARIDADES	63
3.2.1	Conceitos e características	67
3.2.2	Utilidade do big data	73
3.2.3	De onde vêm os dados?	74
3.3	USOS PRÁTICOS DO BIG DATA	79
3.3.1	Big data e o Governo	80
3.3.2	Big data e espionagem	84
4	OS IMPACTOS DO BIG DATA: DA VIGILÂNCIA À PRIVACIDADE DOS DADOS	87
4.1	SOCIEDADE VIGIADA	87
4.1.1	A vigilância distribuída	91
4.2	A BUSCA POR PADRÕES DE COMPORTAMENTO.....	93
4.3	QUEM VIGIA QUEM NAS REDES SOCIAIS.....	98
4.4	DADOS COLETADOS PELAS MÍDIAS E REDES SOCIAIS	100
4.5	AS IMPLICAÇÕES DO BIG DATA NA COLETA DE DADOS PESSOAIS	105
4.5.1	Direito à privacidade dos dados	107
	PARTE II – O CAMPO DO OBJETO DE PESQUISA	112
5	CAMPANHAS ELEITORAIS: CONTEXTUALIZAÇÃO E CARACTERIZAÇÃO	113
5.1	DEFINIÇÃO E EVOLUÇÃO DAS CAMPANHAS ELEITORAIS.....	113
5.2	CAMPANHAS ELEITORAIS: DA MASSIFICAÇÃO PARA A SEGMENTAÇÃO .	119
5.2.1	O uso do marketing na política	123
5.2.2	Do marketing eleitoral para o marketing individual	125
5.3	A INFORMAÇÃO E A PESQUISA NO AMBIENTE ELEITORAL	128
5.3.1	Case Basílio Villani	130
5.3.2	O papel da pesquisa na campanha eleitoral	133

5.3.3 As mídias sociais como estratégias de pesquisa eleitoral.....	136
5.3.4 Cenário das redes sociais nas eleições brasileiras de 2014	138
5.4 CAMPANHAS ELEITORAIS CENTRADAS EM DADOS.....	140
5.4.1 O eleitor é o alvo	142
5.4.2 Microtargeting eleitoral.....	147
5.5 CASE BARACK OBAMA 2012.....	149
5.5.1 Campanha eleitoral OFA	150
5.5.2 Equipe e tecnologia	155
5.5.3 A segmentação da comunicação na campanha.....	159
5.5.4 As redes sociais na campanha	162
PARTE III – A PESQUISA DE CAMPO	165
6 VISÕES SOBRE OS USOS DO BIG DATA EM CAMPANHAS ELEITORAIS BRASILEIRAS	166
6.1 METODOLOGIA DA PESQUISA.....	166
6.1.1 Objetivos da pesquisa	166
6.1.2 Instrumentos e técnicas de pesquisa	167
6.1.3 Coleta de dados e análise dos resultados.....	168
6.1.4 Fases da pesquisa de campo	168
6.2 EMPRESAS BRASILEIRAS DE PESQUISA E DADOS.....	171
6.2.1 Elife.....	171
6.2.2 Ideia Inteligência	174
6.3 CONSIDERAÇÕES DA PESQUISA DE CAMPO	178
6.4 VISÕES DOS ESPECIALISTAS SOBRE BIG DATA	182
6.4.1 Mudanças com o big data	184
6.4.2 O big data em campanhas eleitorais.....	187
6.4.3 Estrutura e especialidades profissionais	191
6.4.4 Usos de dados pessoais	194
6.4.5 Pontos negativos do big data	197
6.5 CONSIDERAÇÕES DA PESQUISA DE CAMPO	200
6.6 USOS DO BIG DATA NA CAMPANHA ELEITORAL DE 2014	201
6.6.1 Cenário eleitoral para governo do Rio de Janeiro	202
6.7 DE DENTRO DO COMITÊ ELEITORAL	203
6.7.1 Inteligência de dados e informações eleitorais.....	207
6.7.2 De onde vêm os dados da campanha	215
6.7.3 Estrutura do trabalho de big data	216
6.7.4 As mídias sociais na campanha	218
6.7.5 Para que serviam os dados e informações na campanha.....	227
6.7.6 As informações na campanha digital.....	228
6.8 CONSIDERAÇÕES SOBRE A PESQUISA DE CAMPO	232
7 CONSIDERAÇÕES FINAIS.....	236
REFERÊNCIAS	243

APÊNDICES	257
APÊNDICE A – ROTEIRO DE ENTREVISTAS COM EMPRESAS	258
APÊNDICE B – ROTEIRO DE ENTREVISTA COM PROFISSIONAIS	259
APENDICE C- ROTEIRO DE ENTREVISTAS COM ESPECIALISTAS	260
APENDICE D- ROTEIRO DE ENTREVISTA COM CIENTISTA POLÍTICA	261
APÊNDICE E - DIÁRIO DE CAMPO	262

1 INTRODUÇÃO

Esta tese tem como foco os usos que campanhas eleitorais contemporâneas têm feito do *big data*, a partir do fenômeno da datificação da informação e da adoção de novas práticas para conquista de eleitores daí derivadas. A escolha por essa temática se justifica pela falta de discussões acadêmicas mais aprofundadas a esse respeito, direcionadas principalmente ao campo¹ da política eleitoral, em especial às campanhas eleitorais brasileiras.

As motivações para a pesquisa levam em considerações os interesses pessoais, profissionais e acadêmicos. O interesse por focar o objeto desta investigação em campanhas eleitorais parte das considerações feitas por Mancini e Swanson (1996), em que as campanhas eleitorais são assuntos difíceis de serem estudados e que o que acontece com elas reflete oportunidades, tradição e personalidades. A partir disso, entende-se que nenhuma campanha é exatamente como outra e, certamente, nenhuma campanha eleitoral é igual de outros países, assim como os métodos e práticas adotadas em cada uma. Dessa forma, acredita-se que práticas de campanhas eleitorais merecem ser examinadas, principalmente no contexto brasileiro.

Esta tese começou a ser desenvolvida no final de 2012, no momento em que o termo *big data* começava a ocupar as pautas de discussões em Congressos e Encontros, e debates acadêmicos se voltavam para a temática. Da mesma forma, reportagens jornalísticas e artigos científicos começaram a ser publicados sobre as possibilidades oferecidas com o uso do *big data* em diversas áreas. Porém, o grande impulso da presente investigação se deu logo após a campanha eleitoral americana para a reeleição do presidente democrata Barack Obama, na disputa eleitoral de 2012. Essa campanha é referenciada por especialistas como um exemplo emblemático dos usos do *big data* na conquista de eleitores, servindo, assim, como referência e incentivo para o desenvolvimento desta tese.

Neste trabalho, busca-se caracterizar *big data* a partir das discussões da sociedade informacional estabelecida por Castells (1999), em que a informação e a tecnologia se colocam como uma discussão central para a sociedade na atualidade. A problematização da pesquisa se dá a partir do excesso de informações que teve o seu aceleração a partir do surgimento da internet e da proliferação das mídias sociais digitais como espaços de circulação de informação e sociabilidade entre as pessoas. Isso tem se refletido significativamente no contexto político eleitoral, em que cada vez mais eleitores se utilizam

¹ Entende-se por campo o espaço em que a pesquisa empírica acontece. Na tese, campo faz referência às campanhas eleitorais.

desses espaços para compartilhar e expressar suas opiniões sobre administrações públicas, políticos, candidatos e processos eleitorais. E que de forma alguma pode ser desconsiderado por profissionais de comunicação e marketing, no momento das construções das estratégias numa disputa eleitoral.

Nesse cenário, a internet e o uso de mídias sociais transferiram o poder de criação de conteúdo para as pessoas, aumentando em muito a geração de informação nesse ambiente. Contudo, especialistas dizem que isso representa apenas uma pequena parte da informação que é gerada diariamente. Hoje, sensores dos mais variados tipos, assim como *smartphones*, contribuem significativamente para o excesso de informação. Segundo estimativas realizadas pela *International Business Machines* (IBM) em 2013, no momento em que essa pesquisa começava a ser delineada, aproximadamente 2,5 bilhões de *gigabytes* de dados eram criados por dia e esse número dobra a cada 40 meses (BRYNJOLFSSON; MCAFEE, 2014). Já para 2015, a estimativa é de que o volume global de dados digitais possa chegar a 8 *zettabytes*. Esse volume exponencial de dados gerados nos últimos 20 anos, particularmente na internet, passou a ser preocupação de muitos especialistas e pesquisadores que começaram a chamar esse fenômeno de “explosão da informação” (LE COADIC, 1996; BURKE, 2012).

No entanto, a preocupação com os dados não é algo recente. Desde os primórdios, a humanidade empregava formas de coletar e analisar dados a fim de extrair informações para diversas finalidades, inclusive como sinônimo de poder, dominação e formação do saber. Mas, nos tempos atuais, a disponibilidade de espaços de compartilhamento de informações na internet e os avanços nas tecnologias computacionais permitem guardar, organizar e analisar dados muito mais facilmente e com frequência muito maior, facilitando e ampliando as ações sobre os indivíduos. Assim, entende-se que esses avanços acabam por contribuir para uma maior vigilância e monitoramento das suas ações e comportamentos, que estão sendo usados para conduzir até mesmo as campanhas eleitorais na atualidade.

Considera-se que esses eventos tenham impulsionado o conceito de *big data*, que até hoje não segue uma definição rigorosa e configura-se como um conceito amplo, caracterizado pelos chamados três Vs, que se referem ao volume, velocidade e variedade de informações (GARTNER, 2013; TAURION, 2013). De maneira geral, *big data* configura-se na convergência de dados de vários ambientes on-line e off-line, que, analisados de forma conjunta, podem levar a descobrir relacionamentos que antes eram impossíveis de se desvendar (TAURION, 2014).

Apesar de ser um conceito relativamente novo no campo científico, já existem casos de utilização dos dados digitais ressignificando redes, teorias e processos nos mais diferentes

domínios do conhecimento (ZUPPO *et al*, 2013). O termo *big data* começou a ser usado na década de 90 para descrever grandes conjuntos de informações que desafiavam os limites computacionais tradicionais de captura, processamento, análise e armazenamento da *National Aeronautics and Space Administration*, a NASA, principal órgão de exploração e pesquisa espacial do mundo, mas para muitos especialistas é um termo que vem do mercado digital configurado por *buzz word*² e que vem sendo chamado no Google Trends, desde 2005, de nova onda do momento. De acordo com Cukier (2010), os efeitos do fenômeno *big data* vêm sendo percebidos na prática em todos os lugares como empresas, governos e ciência. Seu conceito está migrando para todos os campos do conhecimento humano, pois em essência seu avanço é uma continuação da antiga busca da humanidade em medir, registrar e analisar o mundo desde os primórdios da composição das sociedades (SCHÖNBERGER-MAYER; CUKIER, 2013).

No campo da política eleitoral, o *big data* é referenciado por muitos especialistas (MOURA, 2014; PEREIRA, 2014; TAURION, 2014) como uma importante transformação na forma como campanhas eleitorais podem realizar trabalhos em grande escala, que não poderiam ser feitos anteriormente para extrair novas ideias e criar novas formas de valor que alterem principalmente as relações entre candidatos e eleitores. No entanto, as contribuições de *big data* para diversas áreas são possíveis de serem realizadas desde que a matéria-prima informacional, gerada pela explosão dos dados, seja transformada em inteligência, por meio do processo de análise.

A partir desse cenário, a pesquisa levanta as seguintes questões:

(1) em que medida o *big data* tem sido usado em campanhas eleitorais?

(2) que tipos de mudanças podem ser percebidas nas campanhas eleitorais brasileiras com os usos do *big data*?

Para responder aos questionamentos da pesquisa parte-se das hipóteses de que:

(1) o uso do *big data* em campanhas eleitorais, inclusive no Brasil, muda as estruturas e formas de se fazerem campanhas;

(2) o grande volume de informação gerado e compartilhado em ambientes digitais impõe novos olhares e estratégias para as campanhas eleitorais;

(3) o *big data* tende a alterar as relações entre eleitores e candidatos, permitindo que a campanha tenha melhor conhecimento em relação ao comportamento dos eleitores, contribuindo para a segmentação eleitoral.

² *Buzz word* significa um termo da moda, cunhada por especialistas do mercado digital e divulgadas por empresas de marketing.

O objetivo desta tese é contribuir com o avanço das investigações sobre os usos do *big data* em campanhas eleitorais. Sublinha-se ainda que não é o objetivo neste trabalho examinar os impactos do *big data* na decisão do voto do eleitor, nem tampouco demonstrar experimentalmente como se utiliza o *big data*, mas apresentar as oportunidades oferecidas pelos usos do *big data* no contexto eleitoral brasileiro e relatar a partir de pesquisas empíricas como campanhas eleitorais estão trabalhando com esse novo conceito.

A partir do objetivo geral, busca-se cumprir os seguintes objetivos específicos:

(a) observar como o *big data* tem sido usado em campanhas eleitorais, verificando as possibilidades, os limites e riscos desse uso;

(b) identificar que estratégias informacionais e comunicacionais podem ser aplicadas no campo político a partir do uso do *big data*;

Para atingir esses objetivos, a construção desta pesquisa se deu a partir de um arcabouço teórico que vai além da Ciência da Informação (CI), buscando assim uma interdisciplinaridade entre outras áreas do conhecimento, partindo da ideia de que uma única disciplina ou área não pode dar conta de responder aos questionamentos feitos nesta investigação.

Considera-se que uma tese de doutorado é sempre um ponto de partida e nunca o ponto de chegada, porque a investigação não se esgota para o pesquisador no momento da finalização das suas pesquisas. A partir disso, passou-se a construir o caminho para buscar entender o tema central da tese. A escolha da metodologia para avaliar as hipóteses propostas levou em consideração as perspectivas de Booth (2000) de que “o ponto de partida de uma boa pesquisa é sempre o que você não sabe ou entende, mas sente que deve conhecer ou entender” (BOOTH, 2000, p. 51). Assim, esta pesquisa caracteriza-se como um estudo exploratório e de natureza qualitativa (GIL, 2008), porque não tem a intenção de quantificar o objeto estudado e sim situá-lo e caracterizá-lo dentro do campo escolhido para investigação.

Os problemas de pesquisa levaram à definição de uma estratégia de investigação organizada em duas componentes principais: levantamento e revisão bibliográfica; pesquisa empírica. No primeiro momento, percebeu-se a necessidade de conhecimento do contexto em que a problemática se inseria. Nessa fase da pesquisa, procura-se um aprofundamento teórico e análise da relação e influências entre os conceitos de sociedade informacional, explosão da informação, internet, datificação, *big data*, campanhas eleitorais e marketing eleitoral. A pesquisa bibliográfica foi realizada em livros e também envolveu levantamento em bases de dados da área das Ciências Sociais Aplicadas, pesquisas em Portal de Periódicos da Capes,

além de outras bases de dados da internet, que apresenta artigos, teses e dissertações sobre o tema.

A pesquisa empírica envolveu desde o planejamento para a coleta de dados até a análise dos resultados, que foi realizada em três fases. Na primeira fase, foi feita pesquisa em empresas de pesquisa e dados brasileiras, que trabalham com *big data* na área de política eleitoral. A pesquisa foi realizada em 2013 e 2014 com duas empresas, a ELife da cidade de São Paulo e a Ideia Inteligência da cidade do Rio de Janeiro. Na segunda fase, foram realizadas 12 entrevistas com especialistas e profissionais, sendo 10 brasileiros e dois americanos, que tiveram como objetivo buscar visões e expectativas a respeito do *big data* e seu uso em campanhas eleitorais. Foram entrevistados jornalistas, publicitários, professores pesquisadores, cientistas de dados, cientista político, cientista social e profissionais da área de tecnologia da informação e também da área de marketing político. As entrevistas foram realizadas durante os anos de 2013, 2014 e 2015 e foram feitas pessoalmente, por e-mail e por Skype. Na terceira e última fase da pesquisa empírica foi realizado uma observação direta em um comitê de campanha eleitoral, que teve por objetivo entender a dinâmica do *big data* nas estratégias de informação e comunicação no momento da disputa eleitoral. Esse acompanhamento se deu juntamente com a equipe de profissionais de comunicação e marketing eleitoral e aconteceu durante o período do primeiro turno das eleições (de agosto a outubro) de 2014. A campanha pesquisada foi de um candidato que disputava o governo do Estado do Rio de Janeiro.

1.1 A ESTRUTURA DA TESE

O presente trabalho está estruturado da forma que se explica a seguir.

Em (1) **Introdução**. Essa parte estabelece os balizamentos da tese conduzindo o leitor para compreensão do que se buscou obter como resultado da pesquisa e a natureza dos procedimentos seguidos.

Para uma melhor organização da leitura depois da introdução, a tese foi dividida em três partes.

Na **Parte I – O enquadramento teórico e temático** está dividido da seguinte forma:

(2) **A sociedade informacional e a datificação da informação**. Nesta parte, problematiza-se e contextualiza-se a informação e o excesso de informação gerado principalmente com o desenvolvimento da internet, uso das mídias sociais digitais e de dispositivos móveis de comunicação. São apresentadas discussões em torno da importância

dos dados e do fenômeno de datificação da informação a partir das possibilidades de usos estratégicos e de inteligência competitiva de informação.

(3) **Big data: diferentes dimensões conceituais e práticas.** Nesta parte o objetivo é estabelecer uma discussão teórica e conceitual sobre o termo *big data*. São apresentados estudos empreendidos na área de ciências sociais e busca-se contextualizar *big data*, apresentando as suas possibilidades em diferentes áreas.

(4) **Os impactos do big data: da vigilância à privacidade dos dados.** Nesta parte do trabalho são apresentadas discussões sobre as implicações e impactos dos usos de dados pessoais como processo de vigilância e monitoramento em mídias sociais digitais, bem como as regulamentações sobre os usos de dados pessoais. Estabelece-se uma discussão a respeito do uso dos dados dos usuários de redes sociais com propósitos políticos em campanhas eleitorais. Por fim, é ilustrado com exemplos instrumentais da coleta e usos dos dados pelas empresas de mídias sociais.

Na **Parte II – O campo da pesquisa** está dividido da seguinte forma:

(5) **Campanhas eleitorais: contextualização e caracterização.** Procura-se conceituar e caracterizar as campanhas eleitorais a partir de discussões em torno de conceitos de marketing aplicados na política eleitoral; destaca-se a importância das pesquisas políticas tradicionais aliadas às opiniões das redes sociais; estabelece-se uma discussão sobre as campanhas eleitorais que são centradas em dados; são apresentadas técnicas e metodologia focadas em microsegmentação de dados. E por fim, exemplifica-se com a campanha eleitoral americana de Barack Obama de 2012.

Na **Parte III – A pesquisa de campo** está dividida da seguinte forma:

(6) **Visões sobre o uso de big data em campanhas eleitorais brasileiras.** São apresentadas as pesquisas empíricas realizadas em empresas brasileiras de pesquisas e dados, com especialistas de diversas áreas do conhecimento e ainda resultados da observação feita em um comitê de campanha eleitoral, na disputa para o governo do Estado do Rio de Janeiro em 2014.

(7) **Considerações.** Por fim, apresentam-se as considerações finais sobre a temática pesquisada desde a pesquisa bibliográfica até a pesquisa de campo, apontando as impressões sobre a temática, bem como sugere-se novas abordagens e novos estudos para o objeto investigado.

PARTE I – O ENQUADRAMENTO TEÓRICO E TEMÁTICO

2 A SOCIEDADE INFORMACIONAL E A DATIFICAÇÃO DA INFORMAÇÃO

Neste capítulo são apresentados os conceitos que conduzem a tese e o objeto de investigação. As discussões são desenvolvidas a partir das temáticas: sociedade informacional; explosão da informação; informação, dado, conhecimento; internet e datificação. Para sustentar essas discussões buscou-se aportes teóricos na Ciência da Informação (CI) e em outras áreas do conhecimento. Entende-se que a partir dessas discussões exista um reconhecimento que a informação se coloca como centro das atuais mudanças vivenciadas na sociedade em que as tecnologias são usadas para melhorar a geração de conhecimentos e o processamento da informação (CASTELLS, 1999).

A intenção aqui não é esgotar o assunto que envolve essas abordagens, mas a partir disso problematizar o objeto de pesquisa e iniciar o delineamento da temática *big data*, que passa a ser tratada com mais profundidade nos capítulos seguintes.

2.1 O CONTEXTO DA SOCIEDADE INFORMACIONAL

A partir da Revolução Industrial, ocorrida na Inglaterra no século XVIII, o processo de trabalho passou por importantes transformações. Três séculos depois, com o desenvolvimento e difusão das tecnologias da informação e comunicação (TICs), os computadores pessoais e as redes de comunicação globais como a Internet, a humanidade é colocada à frente de uma nova onda de transformações. E nesse cenário a informação se coloca como centro das atenções e vira sinônimo de capital e estratégia de competição. Atualmente, a expressão “tempo real” aparece com frequência dando a ideia da velocidade em que corre a informação por meio dos canais de comunicação distribuídos pelo mundo. A internet surge na década de 1960 nos meios militares pela Agência de Projetos de Pesquisa Avançada do Departamento de Defesa dos Estados Unidos, que tinha como objetivo impedir a tomada de destruição do sistema norte-americano de comunicações pelos soviéticos, em caso de guerra nuclear (CASTELLS, 1999). A partir do seu desenvolvimento, a internet passou a ser considerada a terceira revolução chamada e propagada como a Revolução Tecnológica.

Diversas correntes teóricas (NEGROPONTE, 2006; LOJKINE, 2002; RIFKIN, 2001; CASTELLS, 1999) consideram a geração e a difusão da informação e do conhecimento como principais fontes de valor e poder no século XXI. Independente de como se denomine o atual momento, há um comum reconhecimento entre esses teóricos de que a informação, o conhecimento e as tecnologias se colocam como centros das atuais mudanças paradigmáticas tanto do ponto de vista econômico, político e social (ALBAGLI; MACIEL, 2011). Para Albagli e Maciel (2011, p.18), “ao tempo em que o papel central da mudança tecnológica

nesse processo tem sido ressaltado pelas diferentes linhas de pensamento, há também a preocupação de se evitar o determinismo tecnológico”. *Determinismo tecnológico* é um pensamento que aposta que um artefato, uma tecnologia, um meio, sempre condiciona os modos de percepção, de cognição e, enfim, de comunicação de uma dada pessoa e/ou cultura. Esse termo (*determinismo tecnológico*) parece ter sido proposto pela primeira vez na obra do sociólogo americano Thorstein Veblen (1857-1929). Aplicada à análise da obra de McLuhan (2002), surgiria a interpretação de que este autor pensaria a evolução das culturas como decorrente de uma afetação direta dos modelos de tecnologias que emergem, fazendo com que sua compreensão ficasse reduzida a uma lógica causal, linear e sequencial, na qual a tecnologia, exclusivamente, determinasse os modos do ser humano. Contrário a isso, Castells (1999, p.43) diz que “a tecnologia não determina a sociedade nem a sociedade escreve o curso da transformação tecnológica, entende-se o processo como um complexo padrão interativo”.

Segundo Santaella (2003), em relação aos pensamentos de McLuhan (2002), “considerar que as mediações sociais vêm das mídias em si é incorrer em uma ingenuidade e equívoco epistemológicos básicos, pois a mediação primeira não vem das mídias, mas dos signos, linguagem e pensamentos, que elas veiculam”.

Levando em conta que o poder cada vez mais se define a partir do controle sobre o imaterial e o intangível que compõem informações, conhecimentos, ideias, gostos e desejos de indivíduos e coletivos, da mesma forma sobre os meios por onde estes circulam (ALBAGLI; MACIEL, 2011), é importante considerar os apontamentos feitos por Castells (1999) sobre o papel da tecnologia na sociedade contemporânea. Esse autor caracteriza o período atual como “Revolução das Novas Tecnologias da Informação” ou “Sociedade Informacional”. Segundo Castells (1999), a geração, o processamento e a transmissão da informação tornam-se fontes fundamentais de produtividade e poder, estando a informação no cerne das discussões. Assim, nada é mais significativo desta revolução tecnológica do que a internet, que é tida como símbolo de uma nova sociedade que emerge na cultura do terceiro milênio (CASTELLS, 1999). Nessa perspectiva, a internet não se apresenta como uma simples “tecnologia da comunicação”, mas como uma ferramenta fundamental direcionada à produção e à difusão da informação. Castells (1999) define sociedade informacional dizendo “que o processamento da informação é focalizada na melhoria da tecnologia do processamento da informação como fonte de produtividade, em um círculo virtuoso de interação entre as fontes de conhecimentos tecnológicos e a aplicação da tecnologia para melhorar a geração de conhecimentos e o processamento da informação” (CASTELLS, 1999, p. 54). Em relação a isso, segundo Albagli (2013), mais do que a informação e conhecimento como “coisas”, é o trabalho de

transformar conhecimento em informação e inovação, o que marca a nova era, no que Castells chamada de sociedade informacional. Assim, entende-se que o central é o desenvolvimento e a difusão de tecnologias para agir sobre a informação e não apenas de informação para agir sobre (como insumo para) a tecnologia” (ALBAGLI, 2013, p. 112).

Portanto, a sociedade informacional é uma sociedade marcada pela produção, circulação, armazenamento e controle de uma quantidade massiva de dados. Isso coloca a sociedade diante de um aspecto problemático da cultura do nosso tempo relacionado à questão informacional: o do excesso de informação, ao qual se convencionou chamar também de fenômeno de “explosão da informação”.

2.2 O EXCESSO DE INFORMAÇÃO

O excesso de informação pode ser definido como a condição na qual grande quantidade de informação de baixa qualidade torna difícil a descoberta e a seleção da informação ou conhecimento que buscamos. Atualmente as pessoas devem saber lidar com a crescente complexidade das tecnologias de informação e comunicação digitais. A informação que recebemos vem de fontes e formatos variados, e por vezes fragmentadas em pedaços sem conteúdo e contexto. A preocupação nesse sentido paira mais na forma como iremos gerenciar e usar e aproveitar esse grande volume de informações a nosso favor. Assim, o grande volume de informações geradas por diversas fontes e em vários ambientes, principalmente digitais pode se mostrar como uma grande estratégia para diversas finalidades, seja elas sociais ou políticas. E nesse sentido vale a discussão, lembrando de como teóricos e pesquisadores vem tratando o tema.

A “explosão da informação” é lembrada em discussões teóricas como sendo provocada por diversos fatores históricos, mas que coloca principalmente o advento da eletrônica (LE COADIC, 1996) e a digitalização da vida cotidiana como elementos centrais dessa problemática. Segundo Burke (2012), “explosão” é uma imagem pessimista para o que outrora foi chamado de maneira otimista de progresso ou crescimento do conhecimento. Essa nova metáfora é combinada com dois conceitos: a expansão rápida e a fragmentação.

Para Burke (2012), do ponto de vista de um consumidor sofrendo de “angústia de informação”, a metáfora tradicional de se sentir afogado ou as metáforas novas de “ruído”, “neblina de dados” ou “sobrecarga de informação” parecem mais apropriadas. “Lemos, por exemplo, que um dilúvio de dados ou um tsunami de dados está se quebrando nas praias do mundo civilizado” (BURKE, 2012, p. 311).

Mas cabe lembrar que, na história da humanidade, desde que a escrita tornou possível gravar textos e preservá-los além da capacidade da memória dos indivíduos, iniciou-se um processo de acúmulo de informação. Assim, a “explosão da informação” recebe contribuições iniciais e significativas a partir do desejo desenfreado pela socialização da informação influenciada por Gutenberg, por volta de 1448, a partir da invenção da imprensa. O aperfeiçoamento das técnicas chinesas da prensa de tipos móveis, construída em placas de metal – material mais resistente na época e de possível reutilização - e também a formulação de uma tinta mais aderente viabilizaram a reprodução de obras em grande volume. Entretanto, é perceptível que o feito de Gutenberg culminou em outros efeitos, quando se avaliam as marcas deixadas pela mudança no cenário da época. Uma delas seria o surgimento de um grande volume de obras, somado à inexistência de um sistema eficaz de catalogação, ou seja, havia uma dificuldade em encontrar a informação disponível. Além do intenso fluxo de conteúdos, que tornava inviável a leitura de tantos artigos. De 1453 a 1503 cerca de oito milhões de livros foram publicados, que foi uma produção de conteúdo maior que toda a sociedade havia gerado via escribas nos 1200 anos anteriores. “Mais que uma ordem de livros o que alguns contemporâneos percebiam era uma “desordem de livros” que precisava ser controlada. Este é certamente um problema com que nós também estamos brigando atualmente, nos primórdios da mídia eletrônica” (BURKE, 2002, p. 175).

A “explosão da informação” encontra um precedente em outra expressão, o da “explosão dos documentos”, introduzida por Otlet, no Tratado de Documentação (1934). A obra de Otlet é considerada um dos marcos teóricos do início da Ciência da Informação. Segundo Robredo:

Sua enorme massa, acumulada no passado, cresce a cada dia, a cada hora de unidades novas em número desconcertante com frequência enlouquecedora. Deles, como da língua, pode-se dizer que podem se a pior e o melhor das coisas. Deles, como da água caída do céu, pode-se dizer que podem provocar a enchente ou o dilúvio ou se expandir em rego benéfico (OTLET, 1934 apud ROBREDO, 2003, p.45).

O fenômeno aqui referenciado, ligado ao crescimento exponencial da produção e uso de documentos em especial nas áreas científico-técnicas, envolvia uma diversidade considerável de profissionais, em diferentes contextos. Para Robredo (2003) esse fenômeno não se restringia a contextos e a categorias de profissionais específicos, mas sim a uma vasta área que envolvia todos os que lidavam com a informação. Nesse contexto surgiu a Ciência da Informação, procurando responder a questões relativas à gestão da enorme quantidade de registros informacionais.

Os anos de 1950 caracterizaram-se pelo crescimento da informação científica, sobretudo sob a forma de relatórios técnicos e por um rápido desenvolvimento dos sistemas automáticos de armazenamento e de recuperação da informação. E isso aconteceu nos mais diversos suportes, marcada, nesta época, pela progressiva associação da tecnologia à sua produção, tratamento e difusão (ROBREDO, 2003). Dessa forma, coube à CI o estabelecimento desses estudos. No entanto, como afirma Brookes (1980, p.125), “os problemas básicos da Ciência da Informação não são novos”, as questões relativas à produção, circulação e consumo da informação não se originam com esta disciplina, mas sim com os registros informacionais quando da invenção da escrita. Alguns teóricos como Black, Muddiman e Plant (2007) observam que a sociedade da informação dificilmente pode ser considerada nova e que a maioria das sociedades ‘civilizadas’ foi informacional. Da mesma forma, Headrick (2000, p.7-8) diz que “a Era da Informação não tem começo, porque ela é tão antiga quanto a humanidade”.

É importante lembrar as contribuições que os grandes conflitos mundiais trouxeram para a era da informação. Segundo Burke (2012), a Primeira Guerra Mundial (1914-1918) trouxe muitas consequências para o conhecimento; de ambos os lados do conflito, houve o recrutamento de várias habilidades acadêmicas e dos próprios acadêmicos para auxiliar no esforço de guerra. Entre os exemplos estão os físicos que desenvolveram métodos navais para detectar a presença de submarinos, os psicólogos que faziam as avaliações dos pilotos de guerra. Além disso, dois milhões de soldados americanos passaram por testes de inteligência. Assim, segundo Burke (2012, p. 327) “observou-se em 1920 que “um dos resultados mais impressionantes da guerra é a ênfase que ela deu à importância nacional da ciência e da pesquisa”. Em consequência disso, departamentos, comissões e conselhos de pesquisas foram criados na França, na Inglaterra, na Rússia e nos Estados Unidos, contribuindo para o crescimento do volume de pesquisa. Os serviços de inteligência também se expandiram como o FBI (1908), o MI5 (1909) e o SIS (futuro MI6) em 1912. “Nessa época, os governos também precisavam de mais informações do que o habitual para os esforços de guerra, enquanto o aumento do controle e da centralização enfrentou menos resistência do que provavelmente ocorreria em tempos de paz” (BURKE, 2012, p.327).

Desde então, ao passar das décadas, estudiosos e filósofos refletem sobre o tema e quase sempre existe um incomodo e uma certa ansiedade em relação a quantidade de informação que vem sendo gerada e disponibilizada. Wurman (1991), por exemplo, considera que isso foi uma verdadeira “explosão da não informação”, demarcada pela produção e

inserção cada vez maior de dados, isso porque, para o autor, “informação é aquilo que leva a compreensão” (WURMAN, 1991, p.43).

No cerne dessa discussão, é surpreendente perceber que, como afirma Wurman (1991, p.36), “uma edição do The New York Times em um dia da semana contém mais informação do que o comum dos mortais poderia receber durante toda a vida na Inglaterra do século XVII”. Ainda sobre o problema de excesso de informação, segundo Burke (2012), uma das razões pelas quais os atentados de 11 de setembro, que marcaram a história dos Estados Unidos, não foram detectados de antemão pelas forças de segurança do país, apesar dos alertas, foi porque esses alertas se perderam na enxurrada de dados (BURKE, 2012, p.184).

Ainda sobre o excesso de informação, é importante lembrar que, por volta de 1940, o escritor americano Arthur Fremont Rider publicou em seu livro “*The Scholar and the Future of the Research Library: a Problem and Its Solution*” uma pesquisa na qual estimou que as bibliotecas americanas dobrariam de tamanho a cada 16 anos e, continuando nessa taxa de crescimento, a biblioteca situada na cidade de Yale teria aproximadamente duzentos milhões de volumes armazenados até o ano de 2040. Esse volume ocuparia, então, cerca de dez mil quilômetros em prateleiras. Para servir como referência, a distância entre os extremos norte e sul do Brasil não chega a cinco mil quilômetros. De fato, a previsão de Arthur Fremont só não acontecerá devido à digitalização de livros e documentos, recurso não existente na época da pesquisa. Caso contrário, devido ao elevado aumento das publicações, essa previsão seria superada bem antes de 2040.

Por outro lado, o que era considerado demasiado antes da era da internet começou a mudar e o processo foi ganhando velocidade na era da informação digital, como se observa no próximo item deste trabalho.

2.2.1 O acúmulo de dados na era digital

Em um discurso realizado em 1990, Neil Postman, transmitiu suas ideias em relação à tecnologia da computação e à informação. Em suas palavras, Postman é categórico em afirmar que a informação se virou contra nós. Para ele havia um tempo em que ela nos ajudava a resolver questões específicas, porém, após Gutenberg e a explosão da informação, o caos foi trazido até os tempos atuais.

Para alguns estudiosos, a chegada do computador no meio social e a evolução da tecnologia, principalmente a móvel, contribuíram na última década para o acúmulo dos dados. O mundo não apenas está mais cheio de informação, como também a informação está se acumulando com mais rapidez. (SCHÖNBERGER-MAYER; CUKIER, 2013). Assim, a

verdadeira revolução na era da informação não está nas máquinas que calculam os dados, e sim nos dados em si e na maneira como a sociedade os usa; a preocupação se volta para a gerência do exponencial volume de dados.

Segundo Marteleto (2010), a internet, chamada “rede das redes”, caracteriza-se por dois aspectos principais. “Primeiro, é um grande acervo de dados e de informações aberto a múltiplas escritas, consultas, leituras, usos e apropriações. Segundo, é uma arena ampliada geograficamente e socialmente para interação, comunicação e sociabilidade” (MARTELETO, 2010, p. 32).

A expansão da rede foi desencadeada pelo aumento exponencial do número de usuários conectados, bem como pelo desenvolvimento das tecnologias de informação e comunicação. “Esse crescimento da web foi favorecido pelo avanço de equipamentos tais como: satélites, roteadores, servidores, *switches*, *backbones* entre outros” (SCHONS, 2007, p.2). Em *The language of the new media*, Manovich (2000) sublinha a paridade entre a informação e a nova tecnologia. “Para melhor ou para pior, o acesso à informação tem se tornado uma atividade chave da era computacional” (MANOVICH, 2000, p.217). Todavia, isso não quer dizer que não havia manipulação de informação antes da emergência da nova mídia, mas sim que foi através dela que o seu acesso pôde ser intensificado e melhorado, levando-se menos tempo para todo o processo. Esse autor afirma ainda que o elemento essencial do meio digital é o dado, sendo que este é usado, manipulado, acessado, armazenado, por dois tipos de uso contidos na nova forma midiática: o *database* e o *navigable space* (MANOVICH, 2000). O primeiro é uma forma de constituir uma coleção de dados a serem armazenados e manipulados com facilidade e rapidez; e o segundo seria a Internet propriamente dita, cuja característica primordial é a facilidade de acesso a dados. Por esse motivo, os dados passaram a ser um importante fator na análise dos fenômenos ligados tanto à tecnologia computacional em geral quanto ao campo relativo ao uso da rede digital.

Se, por um lado, a facilidade de acesso e publicação de documentos na rede desencadeou novas possibilidades quanto à construção de um grande acervo informacional, por outro também resultou em problemas de excesso, como aponta Schons (2007):

Tal fenômeno a princípio é positivo, mas no decorrer do tempo esse processo originou um grande problema na rede pelo fato de conduzir e fomentar o processo de explosão informacional gerando grandes conflitos, uma vez que o excesso de informações parece ser pior que a sua falta (SCHONS, 2007, p 04).

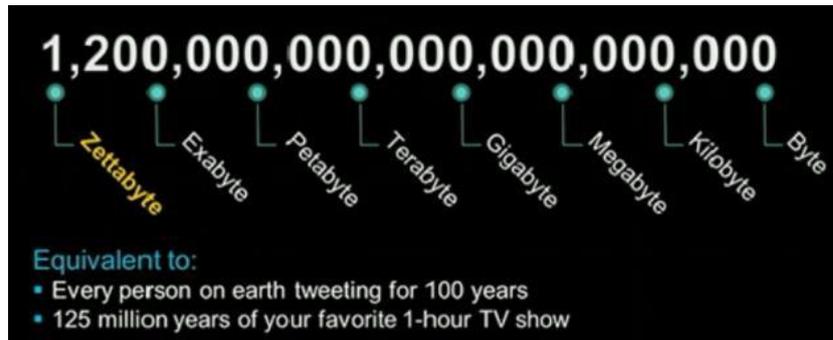
Para o autor, o excesso de informações na rede excede a capacidade humana de explorá-las, ocasionando assim uma grande tarefa temporal para o indivíduo acessar e processar a informação que deseja.

Da mesma forma, Lévy (1999) relaciona a existência atual de um “dilúvio” das informações caracterizado pela proliferação destas por meio do avanço das tecnologias ligadas aos bancos de dados, hipertextos e redes. Há muitos dados a serem manipulados. Esta seria a essência do “dilúvio”. No “dilúvio informacional”, as informações brotam de todos os lados. Lévy observa que o “dilúvio informacional”, na Internet, é fluido, virtual e ao mesmo tempo reunido e disperso. Para ele, “essa biblioteca de Babel não pode ser queimada. [...] As águas deste dilúvio não apagarão os signos gravados: são inundações de signos” (LÉVY, 1999, p.16).

Terra (2000) diz que o excesso de informação está associado à perda de controle sobre a informação e à imobilidade em usá-la. Da mesma forma, Schons (2007) observa que hoje as atenções estão voltadas para o desenvolvimento de tecnologias visando à busca e à recuperação eficiente das informações, já que as dificuldades de selecionar e escolher informações relevantes em meio a um espaço não estruturado crescem na mesma proporção que o aparecimento delas. Para Lima Júnior (2011), a capacidade e facilidade em reunir e armazenar informações em um banco de dados são desenvolvidas e propagadas para facilitar o trabalho do consumidor.

Já para Taurion (2014), a capacidade de armazenamento de dados está duplicando a cada 12 meses. O autor ilustra essa afirmativa com o estudo, publicado na revista *Science*, que mostra que, até 2007, a quantidade de dados armazenados mundialmente foi de 295 *exabytes*. Isso equivale a cerca de 1,2 bilhão de discos rígidos da época. Taurion observa que o estudo ainda mostra que, em 2000, 75% da informação eram guardadas em formatos analógicos, como o vídeo cassete. Já em 2007, 94% dos dados eram digitais. Essa pesquisa aponta a chegada da era digital em 2002, primeiro ano em que a capacidade de armazenamento digital ultrapassou a analógica. As conclusões de um dos pesquisadores são usadas para ilustrar essa pesquisa no trabalho de Taurion (2014): “se fôssemos pegar todas essas informações e armazená-las em livros, poderíamos cobrir toda a área dos EUA ou da China em três camadas. Se a mesma informação fosse armazenada digitalmente em CDs, a pilha de discos criada poderia chegar à Lua”. Os resultados do estudo também mostraram que a humanidade transmite diariamente cerca de dois *zettabytes* de dados. Isso equivale a 175 jornais por pessoa, por dia (TAURION, 2014). A Figura 1 ilustra a contagem de *bytes*.

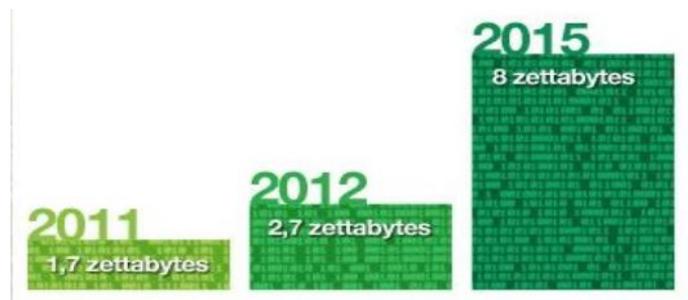
Figura 1: Contagem de bytes



Fonte: Magalhães (2013)

Assim, um *zettabyte*, por exemplo, equivale a todas as pessoas do mundo tuitando por 100 anos, ou assistindo a um programa de televisão por uma hora durante 125 milhões de anos. Essa comparação dá a dimensão do volume de dados que a sociedade está gerando nos últimos tempos. Ainda segundo os estudos da IDC (*Internet Data Center*)³, o volume de informação digital de 2011 para 2015 totalizou 8 *zettabytes*, como mostra a Figura 2.

Figura 2: Volume de informação digital



Fonte: IDC (2015)

A IDC ainda estima que 90% das informações digitais que estão sendo criadas são não estruturadas, ou seja, em diversos formatos como textos, vídeos, áudio e imagens, vindas de mídias sociais digitais. São informações que não estão armazenadas nos sistemas corporativos controlados pela área de Tecnologia da Informação (TI). Elas não estão guardadas em bancos de dados, aplicativos de gestão empresarial; estão espalhadas em diversos formatos na Internet, principalmente nas mídias sociais. E isso tem se tornado uma preocupação constante das áreas que buscam conhecer o comportamento humano por meio dessas informações

³ IDC é a empresa líder em inteligência de mercado e consultoria nas indústrias de tecnologia da informação, telecomunicações e mercados de consumo em massa de tecnologia. Analisa e prediz as tendências tecnológicas para que os profissionais, investidores e executivos possam tomar decisões de compra e negócios nestes setores.

desestruturadas e que cada vez mais carregam informações valiosas sobre a sociedade. A partir disso, outros institutos de pesquisas fazem estimativas com relação ao volume de informações futuras. A IBM, por exemplo, prevê que o volume de informações em 2015 será equivalente a 1850 *exabytes* (mais de 2 milhões de *gigabytes*), ou o equivalente a 92,5 milhões de bibliotecas do Congresso Nacional Americano, que possui mais de 155 milhões de itens.

Dessa forma, não há dúvidas de que a era digital contribuiu significativamente para o crescimento desenfreado de informações. Para isso, basta verificar pesquisas que são feitas anualmente sobre a produção e circulação de informações nos meios digitais. Segundo estudos do EMC⁴ (empresa líder do mercado internacional de armazenamento de dados), em 2014, já existiam disponíveis no mundo quase 1 septilhão de bits de informação – ou o número 1 seguido de 24 zeros – total similar ao de estrelas conhecidas no céu. As informações disponíveis em formato digital equivaleriam a uma pilha de iPads Air (os mais finos) de, nada menos, que 256 mil quilômetros, ou dois terços da distância do nosso planeta até o seu satélite. No fim da década, no entanto, seriam seis pilhas e meia: 1,6 milhão de quilômetros. Se todas essas informações fossem divididas pela população conectada em 2020, haveria quase 6 mil gigabytes de carga por pessoa. A EMC ainda prevê que até 2020 o número de dados armazenados em computadores, servidores, celulares, smartphones e *tablets* seja multiplicado por seis.

Especialistas também alertam sobre o crescimento da quantidade de dados digitais que vem sendo gerado. Letouzé (2012) já indicava que a quantidade de dados digitais disponíveis em nível global cresceu de 150 exabytes em 2005 para 1.200 exabytes em 2010, alertando para um aumento de 44 vezes entre 2007 e 2020. Para Rappouteur (2010), a explosão das redes móveis, a computação em nuvem e as novas tecnologias deram origem a um número incompreensivelmente grande de informações.

Em estudos, Pimenta (2013) ressalta as pesquisas realizadas pelo Centro de Estudos sobre as Tecnologias da Informação e da Comunicação em 2013, em que contabilizou pouco mais de três milhões e duzentos mil domínios com o registro “.br”. Segundo o autor:

De fato, há um número mais extenso de endereços eletrônicos, de dados e informações produzidas cotidianamente por todos nós. Informações em múltiplos níveis. Da internet à intranet, de compartilhamentos nas redes sociais às mensagens de texto via celulares, câmeras digitais, tablets e smartphones. A cada clique, ou hit, que realizamos compartilhamos informações via depoimentos, entrevistas, imagens,

⁴ Informações retiradas da reportagem do O Globo on-line. Disponível em: <<http://oglobo.globo.com/sociedade/tecnologia/estudo-da-emc-preve-que-volume-de-dados-virtuais-armazenados-sera-seis-vezes-maior-em-2020-12147682>>

vídeos e documentos digitalizados. Deixamos, ademais, uma espécie de “rastros”, ou traços das atividades realizadas. Assim, temos produzido em escala exponencial cada vez mais dados por meio do uso do ciberespaço (PIMENTA, 2013, p.4).

Pelo prisma comercial, Taurion observa que:

No dia a dia, a sociedade gera cerca de 2,5 quintilhões de bytes de informações sobre as suas operações comerciais e financeiras, bem como sobre clientes e fornecedores, nas informações que circulam nas mídias sociais e dispositivos móveis como as geradas pelo número cada vez maior de sensores e de outros equipamentos embutidos no mundo físico, como rodovias, automóveis e aeronaves. Um único segundo de vídeo em alta definição gera 2 mil vezes mais bytes que uma página de texto (TAURION, 2014, p.120).

No entanto, a partir dessa ótica histórica, é possível ressaltar que o excesso de informação sempre existiu. Não é um tema novo como observa Johnson quando diz que:

Mesmo assim, não só culpamos a informação por nossos problemas, nós somos arrogantes a respeito. Mais preocupante do que a nossa personificação de informação é a presunção de que o conceito de excesso de informação é algo novo, específico do nosso tempo (JOHNSON, 2011, p. 744).

Nesse cenário, capturar, manusear e analisar esse imenso volume de dados e informações se colocam como grandes desafios. A preocupação que se instala na atualidade é de como se fará para dar conta de administrar e tirar valor desse excesso de informação. Que metodologias, ferramentas ou tecnologias poderão ajudar na solução desse problema? Essa é uma questão que será debatida no próximo capítulo quando será tratada a temática *big data*. No item a seguir apresento os conceitos sobre informação, dado e conhecimento e estabeleço relações entre esses elementos que norteiam o objeto desta investigação.

2.3 A INFORMAÇÃO COMO CONCEITO INTERDISCIPLINAR E MODERNO

Na sociedade informacional (CASTELLS, 1999), conforme retratada anteriormente, coloca-se a informação como matéria-prima e como fatos de produção e desenvolvimento humano e social. A informação assume esse estatuto devido essencialmente à sua natureza digital (CAPURRO, 2007) e que vem sendo motivo de preocupações nos dias de hoje.

Assim, compartilha-se das concepções de Capurro⁵ (2003) que entende que, em um contexto de novos tempos, é preciso entender a informação como um conceito interdisciplinar, multidisciplinar e transdisciplinar, por meio de abordagens que defendem uma visão mais ampla e menos restritiva do objeto das ciências. Portanto, segue-se uma corrente que se apropria de uma abordagem eclética, flexível e sem reducionismos. Esse

⁵ Capurro não acredita numa intenção metafísica de explicar “tudo” e de fazer da informação uma espécie de fluído que se transforma passando de um nível para outro. A abordagem aqui se convencionou a chamar de Trilema de Capurro (2003).

pensamento de Capurro (2003) leva o entendimento do conceito de informação sem restrições próprias de domínios de saber, onde se possam ver as diferenças e semelhanças em várias ciências, em outros contextos, seja cultural, social ou político e em outras épocas e culturas, num enfoque epistemológico que mostre os diferentes campos de aplicação desses conceitos.

Considera-se aqui a informação no sentido que possa contribuir para o conhecimento de determinado fenômeno ou situação, e que está registrada e circula por diferentes meios, sendo socializada pelos indivíduos e que dela possam se tirar respostas para diferentes finalidades. Nesse sentido, Choo (2003, p.119) diz que a “informação precisa possuir algumas características como, por exemplo: responder a uma questão, solucionar um problema, subsidiar uma decisão, auxiliar em uma negociação ou dar sentido a uma situação”. Também existem autores que defendem que a informação se configura em dados organizados ou matéria informacional potencialmente significativa, que serve para a tomada de decisão (URDANETA, 1992; MIRANDA, 1999).

É importante considerar que quando se usa o termo informação em Ciência da Informação (CI) deve sempre lembrar que “informação é o que é informativo para determinada pessoa” (CAPURRO; HJORLAND, 2007, p 154). Destaca-se também o processo de cognição do indivíduo como elemento importante, fazendo com que o que é informativo dependa das suas necessidades e habilidades interpretativas, muitas vezes compartilhadas com os membros de uma mesma comunidade de discurso. No contexto cognitivo vale destacar as considerações de Brookes (1980), que defende que o homem não pode ser excluído do processo da informação. Da mesma forma, para Jeanneret (2005) a informação não é uma pura função técnica, ela se constitui num gesto social que outorga, segundo as épocas e os contextos, um poder de fato a uns ou a outros.

Nesse sentido, este trabalho também atenta-se para a informação sob o prisma da discussão estratégica. Nesse contexto, Choo (2003) observa que quem lidava ou lida melhor com a informação e dela sabe tirar proveito consegue levar vantagem, através de uma melhor tomada de decisões, com inteligência, criatividade e, ocasionalmente, esperteza. Corroborar-se a afirmação de Beal (2008) de que a informação é um fator essencial ao desenvolvimento organizacional e também humano.

A informação é um elemento essencial para a criação, implementação e avaliação de qualquer estratégia. Sem o acesso a informações adequadas a respeito das variáveis internas e do ambiente onde a organização se insere, os responsáveis pela elaboração da estratégia não têm como identificar os pontos fortes e fracos, as ameaças e oportunidades, os valores corporativos e toda variedade de fatores que devem ser considerados na identificação de alternativas e na tomada de decisões estratégicas (BEAL, 2008, p.75).

No entanto, é preciso entender que a informação dispersa não constitui inteligência. A partir da estruturação da informação é que a inteligência passa a existir. Weick *apud* Nonaka; Takeuchi (1997, p. 635) desenvolveu a teoria da inteligibilidade, segundo a qual a realidade é uma realização contínua que surge dos esforços de criar a ordem e tentar entender o que ocorre. A inteligibilidade enfatiza que as pessoas tentam tornar as coisas racionalmente responsáveis para si mesmas e para os outros, ou seja, a premissa básica para se fazer inteligência é agregar valor à informação, é estruturá-la de modo que passe a ter um valor, uma importância contextual.

Assim, é importante conceituar o que vem a ser inteligência competitiva. Entende-se como inteligência competitiva a atividade de coletar, analisar e aplicar informações relativas às capacidades, vulnerabilidades e intenções dos concorrentes, ao mesmo tempo monitorando o ambiente competitivo em geral. A inteligência competitiva trata de:

"[...] ler as entrelinhas do site do concorrente ou mesmo das informações tornadas públicas. Envolve também as conversas com os colegas em eventos, e, sobretudo, saber para onde olhar, o que perguntar e o que fazer com os dados que se descobrem" (STAUFFER, 2004, p. 5).

Um sistema de inteligência competitiva busca transformar dados em informação e estes em inteligência a partir dos conhecimentos adquiridos. "Inteligência competitiva é o processo sistemático que transforma pedaços e partes aleatórias de dados em conhecimento estratégico" (TYSON, 2002, p. 1-3). Tarapanoff (2001), diz que a inteligência competitiva é uma metodologia para monitoramento das informações de um ambiente que permitirá, ao tomador de decisão, adquirir conhecimento estratégico e entender o comportamento de seus concorrentes-chave.

A partir das discussões apresentadas, este trabalho entende a informação como algo em movimento e que permite a geração de diferentes conhecimentos individual e coletivo. Além disso, é preciso usar a inteligência competitiva para a tomada de decisão, diante do grande volume de informações disponíveis e dispersos nos diferentes ambientes on-line e off-line.

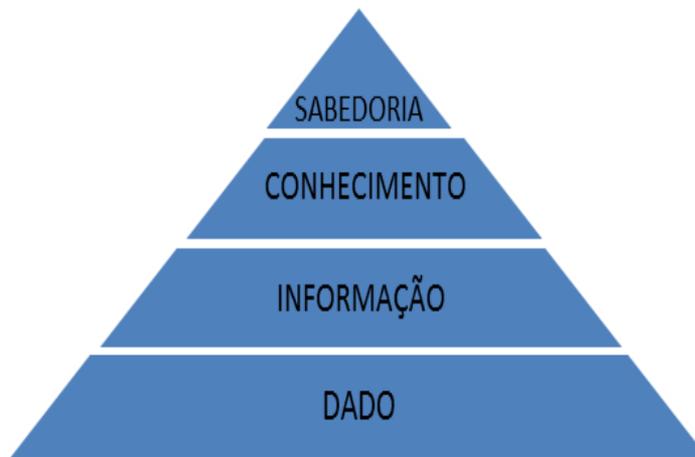
2.3.1 As relações entre informação, dado e conhecimento

É comum relacionar a informação à definição de mensagem, conteúdo, dados, sendo às vezes confundida até mesmo com conhecimento. Mas uma coisa é preciso ressaltar: informação não é a mesma coisa que dado e nem é conhecimento.

Os três conceitos (dado, informação e conhecimento) são muitas vezes considerados elementos básicos na construção de campo de estudos da informação. Dessa forma, a

formulação das suas definições é também entendida como crucial. É comum na literatura de Ciência da Informação encontrar significados diversificados para cada conceito. Por outro lado, os três conceitos estão inter-relacionados e seus significados geram discussões entre os teóricos. Alguns pesquisadores consideram que dados, informações e conhecimento constituem uma hierarquia lógica, como demonstra a Figura 3.

Figura 3: Pirâmide do Conhecimento



Fonte: Adaptado de OGC (2007)

Na pirâmide (Figura 3), o dado e informação são constituídos como o alicerce para a construção do conhecimento e da sabedoria. Para Davenport e Prusak (1998), os dados, informações e conhecimento estão intimamente relacionados formando uma hierarquia entre eles. Esses autores explicam que o conhecimento é decorrente da informação, que, por sua vez, deriva de um conjunto de dados.

Por outro lado, Barreto (2002) ilustra essa relação com base em outra pirâmide que busca qualificar para se compreender, através da imagem, pela largura de cada degrau sobreposto, com a base larga e o ápice afunilado, como mostra a Figura 4.

Figura 4: Pirâmide de Fluxos e Estoques



Fonte: Barreto (2002)

Na Figura 4 a base da pirâmide mostra um estoque subjetivo de ideias, fatos e sensibilidade no indivíduo; o próximo degrau integra a informação como estoque num degrau menor; ela passa os dois degraus como fluxo de conhecimento e da inteligência, até chegar ao superior; o ápice, o saber como menor degrau atingido. Dessa forma, a informação, além de comparada a um estoque na pirâmide de Barreto (2002), é também identificada em um fluxo “segmento, seqüência, sucessão, de eventos dinamicamente produzidos, que determinam o encadeamento ou a vicissitude dos acontecimentos relacionados com as práticas da informação.” (BARRETO, 2002, p. 67). Nos dois momentos, há que se destacar o envolvimento do sujeito com a informação, portanto a informação pode estar além dos suportes.

Partindo de outras conceituações, pode-se dizer que “dados se referem a algo que permite ser registrado, analisado e reorganizado” (SCHÖNBERGER-MAYER; CUKIER, 2013, p. 54). Assim, os dados podem ser também registros guardados num computador que através do processamento da informação se transformam em “informação como-coisa” (BUCKLAND, 1991; LE COADIC, 1996). Ainda sob essa mesma ótica, o dado é entendido como uma seqüência de símbolos quantificados ou quantificáveis. Textos, imagens, sons, animações podem ser transformados em dados e, assim, ser quantificados (MIRANDA, 1999). Por esse prisma, o dado é necessariamente uma entidade matemática e, dessa forma,

puramente *sintática*. Isso significa que os dados podem ser totalmente descritos com representações formais e estruturais.

Já outros autores consideram importante estabelecer distinção entre os conceitos de dado e informação, porque permite definir separadamente a necessidade das bases de dados e as necessidades de informação para gerenciar esses dados (MURDICK; MUNSON, 1988, p. 147).

A informação (na ciência do comportamento) é um signo ou conjunto de signos que impulsionam uma ação. Se distingue de dados porque dados não são estímulos de ação, mas simplesmente cadeias de caracteres ou padrões sem interpretação. Sendo assim, a evolução de dados para informação se mostra uma etapa muito importante no caminho para o objetivo de geração de inteligência. É o início do processo onde não pode haver erros nos dados coletados nem uma má interpretação destes dados, pois o mesmo acarretaria uma tomada de decisão igualmente equivocada ao final do processo (MURDICK; MUNSON, 1988, p.27)

Entende-se que dados não possuem significados para o tomador de decisão. Somente após o tratamento destes dados estes poderão ser transformados em informação e disponibilizados no momento e forma adequada para serem utilizados com eficiência pelo usuário. Considera-se também o conceito de banco de dados ou bases de dados como sendo o armazenamento individual ou combinatório de qualquer tipo de dado ou objeto digital, desde documentos textuais, a fotografias, clips, sequências de áudio, imagens, estáticas, em movimento, ou ainda animações, mapas, gráficos, entre outros, que podem ser navegados, acessados de modos variados (BARBOSA, 2004).

Pelo que se vê, as relações entre dados, informações e conhecimento são bastante estreitas, embora designem diferentes estados. Essa relação de proximidade e distinção é pertinentemente apresentada por Davis e Olson (1987). “Em síntese, os termos dados e informação com frequência são utilizados em formas intercambiáveis, porém a distinção consiste no fato que os dados elementares são a matéria-prima para prover a informação” (DAVIS; OLSON, 1987, p. 209).

Já os pesquisadores e autores especializados em inovação de processos de negócios e gestão do conhecimento Thomas Davenport e Laurence Prusak (2003) defendem uma abordagem ecológica para o gerenciamento da informação. Para eles a “velha distinção entre dados, informação e conhecimento [...] é nitidamente imprecisa”. Segundo esses autores a diferença entre dado e informação reside exatamente no conceito de mensagem inerente à informação, estando presentes os papéis de emissores e receptores desta mensagem. A informação, assim, teria como finalidade exercer algum impacto sobre o julgamento do

receptor, ou seja, “informação são os dados que fazem a diferença” (DAVENPORT; PRUSAK, 2003, p.4).

Já Lévy (1996) aborda a relação entre informação, conhecimento (e saber) como um ato de criação.

[...] Quando utilizo a informação, ou seja, quando a interpreto, ligo-a outras informações para fazer sentido, ou, quando me sirvo dela para tomar uma decisão, atualizo-a. Efetuo, portanto um ato criativo, produtivo. O conhecimento, por sua vez, é o fruto de uma aprendizagem, ou seja, o resultado de uma virtualização da experiência imediata. Em sentido inverso, esse conhecimento pode ser aplicado, ou melhor, ser atualizado em situações diferentes daquelas da aprendizagem inicial. Toda aplicação efetiva de um saber é uma resolução inventiva de um problema, uma pequena criação (LEVY, 1996, p. 58-59).

Portanto, para localizar a informação relevante na imensidão de dados disponibilizados em qualquer ambiente (on-line ou off-line), é necessário ter um conhecimento básico do que se deseja, assim como conhecer estratégias de busca que possibilitem a identificação de fontes confiáveis. O acesso a grandes quantidades de informação não necessariamente garante a possibilidade de transformar em conhecimento.

Nesse sentido, o que seria o conhecimento? Na definição clássica de conhecimento, originada em Platão, consiste de crença verdadeira e justificada. Aristóteles divide o conhecimento em três áreas: científica, prática e técnica. Além dos conceitos aristotélico e platônico, o conhecimento pode ser classificado em uma série de designações. Burke (2012), por exemplo, lança mão de uma metáfora de Claude Levi-Straus para discutir a relação entre conhecimento e informação, pela qual “poderíamos pensar na informação como o cru, enquanto o conhecimento seria o cozido” (BURKE, 2012, p.14), ou no sentido de que a informação apresenta dados menos “processados”, enquanto o conhecimento passa por várias etapas de cozimento/processamento como verificação, crítica, mediação, comparação e sistematização.

Para Calvo conhecimento deve estar relacionado ao estado cognitivo humano.

A produção em massa de dados e fatos e números não significa nada se não houver produção do conhecimento. O conhecimento é criado por pessoas, individualmente, a partir de experiências individuais, eliminando-se o “ruído” e só então é socializado, construindo-se redes de conhecimento (individuais e sociais), e que estamos continuamente atualizando (CALVO, 2004, p.1).

Da mesma forma Le Coadic (1996) diz que o “um conhecimento (um saber) é o resultado do ato de conhecer, ato pelo qual o espírito apreende um objeto” (LE COADIC, 1996, p.4). Na perspectiva de Brookes (1980) a estrutura de conhecimento pode ser objetiva – expressão do pensamento depositado em registros físicos como livros, documentos – e subjetiva- ocorrência no espaço individual privado de cada pessoa. A ideia básica do autor é

de que o conhecimento se dá quando a informação é percebida e aceita, sendo toda alteração provocada no estoque mental de saber do indivíduo, oriunda da interação com estruturas de informação. Assim, também para Barreto (1997), o conhecimento, pode ser visto como toda alteração provocada no estado cognitivo do indivíduo, isto é, no seu acervo mental de saber acumulado, proveniente de uma interação positiva com uma estrutura de informação, o que nos levaria à concepção de ideias de informação como processo e um meio para se atingir determinado fim. “É também desenvolvimento, de uma forma ampla e geral, como um acréscimo de bem estar; um novo estágio de qualidade de convivência que seria alcançado através da informação” (BARRETO, 1997)⁶.

Já Carvalho e Tavares (2001) defendem que o conhecimento precisa partir de informações contextualizadas.

Sabemos [...] que o conhecimento é formado a partir da informação. Entretanto, existe uma variável bastante sutil que, por muitas vezes, esquecemos e que dá ao conhecimento uma característica de fruto ou de produto de uma associação de informações, dentro de um determinado contexto considerado. Isso significa que, assim como informação não é um coletivo de dados, conhecimento também não é um coletivo de informações. Dados carecem de um contexto para se transformar em informação. Da mesma forma, informações carecem de contexto para gerarem conhecimento. E carecem de alguma coisa mais: do talento humano. Uma pessoa pode ter acesso a inúmeras informações, mas se não souber colocá-las em um contexto, interpretá-la diante do mesmo e traduzi-las em uma realidade, não estará gerando conhecimento (CARVALHO; TAVARES, 2001, p.46).

Para Burke (2012) as tradições do ou dos conhecimentos devem ser concebidas no plural e não no singular. Ele cita as concepções feitas pelo filósofo Michel Foucault nos anos de 1970. Foucault (2005, p. 110) define saber como “o conjunto dos elementos (objetos, tipos de formulações, conceitos e escolhas teóricas) formados a partir de uma só e mesma positividade, no campo de uma formação discursiva unitária”. E o conhecimento é por ele definido como “a estranha instauração do formal em uma ordem sucessiva que é a das gêneses psicológicas ou históricas; mas é o ordenamento do empírico por uma forma que lhe impõe sua teleologia” (FOUCAULT, 2005, p.117).

O conhecimento, portanto, impõe-se ao real e visa ordená-lo de fora para dentro. Quem conhece algo está colocando este algo em uma ordem que é externa a ela. Para Foucault (2005) conhecer é uma demonstração de poder. Quem conhece detém poder em relação ao que é conhecido e busca, ainda, compreender o poder do qual dispõe. Dessa forma, Foucault (2005) também considera que o conhecimento deve propor novos caminhos. Assim,

⁶ Citação retirada do artigo A oferta e a demanda da informação: condições técnicas, econômicas e políticas. Disponível em: <http://www.scielo.br/pdf/ci/v28n2/28n2a09.pdf>. Acesso em: 12 de março de 2014.

Burke (2012) divide os conhecimentos em explícitos e implícitos (ou tácitos), puros e aplicados, locais e universais. Albagli (2005) sobre o conhecimento tácito diz que:

O conhecimento tácito é justamente aquele que não foi documentado e tornado explícito por quem o usa ou detém; ele deriva da experimentação e da difusão de novas práticas pela interação local. Já o conhecimento codificado é mais facilmente descrito e replicado de um lugar a outro (ALBAGLI, 2005, p. 4).

Para a autora um dos supostos fundamentais é que parte importante da base de conhecimentos é tácita. Ela considera aqueles que são baseados em crenças, valores, *know-how* e habilidades de cada indivíduo e organização, e provêm do aprender fazendo, usando e interagindo. Assim, para transformar informações em conhecimento, o esforço exercido é o de relacioná-las com as experiências das pessoas (ALBAGLI, 2005).

Em contrapartida o conhecimento explícito é geralmente qualificado de “objetivo” e mais simples de ser codificado, ou seja, formalizado com palavras, números e fórmulas, para ser transmitido rapidamente e em grande escala (NONAKA; TAKEUCHI, 1997).

Já no contexto dos grandes volumes de informação e dados, a relação entre dado, informação e conhecimento é entendida também sob a perspectiva de uma hierarquia reversa (TUOMI, 1999) como é representado na Figura 5.

Figura 5: Hierarquia reversa do conhecimento



Fonte⁷: Disponível em: <http://dcm.ffclrp.usp.br/~augusto/teaching/ami/AM-I-KDD-DM.pdf>.

⁷ Disponível em: <http://dcm.ffclrp.usp.br/~augusto/teaching/ami/AM-I-KDD-DM.pdf>. Acesso em: 10 de outubro de 2014.

Nesse caso, o conhecimento previamente existente uma vez articulado possibilita a compreensão e estruturação da informação que, por sua vez, possibilita a verificação e a existência dos fatos que emerge enfim os dados. Assim, permite-se tirar o dado da sua forma estática, para um estado dinâmico, possibilitando o cruzamento e relacionamento de informações necessárias para a geração do conhecimento. Da mesma forma o conhecimento ao ser explicitado, difundido através de meios físicos ou virtuais, torna-se uma informação ou um conjunto delas, que, ao serem desmembradas, tornam-se um conjunto de dados. Nesse sentido, Davenport e Prusak (1998) afirmam que o conhecimento está relacionado ao uso inteligente da informação, podendo ser avaliado pela qualidade das ações ou das decisões a que leva.

No item a seguir, busca-se a partir de um pequeno resgate histórico situar como a coleta de dados e informações sempre foi prática ao longo dos tempos, na constante busca por conhecimentos, e que acabou por fazer surgir o conceito de datificação.

2.4 O INÍCIO DA DATIFICAÇÃO

A história mostra que não é de hoje que os dados têm se mostrado importantes objetos de coleta, armazenamento e análise, e que a humanidade tem procurado retirar sentido a partir de correlação das informações, seja de forma informal ou com métodos mais científicos.

Schönberger- Mayer e Cukier (2013), numa breve discussão histórica, observam que a escrita foi desenvolvida na antiga Mesopotâmia porque os burocratas queriam um instrumento eficiente para registrar e manter o controle de informações. Desde os tempos bíblicos, os governos organizam censos para reunir dados sobre os cidadãos. O registro de dados mais antigos remonta a 8000 a. C., quando comerciantes sumérios usaram continhas de barro para denotar os bens comercializados. Segundo os autores os seres humanos há muito tempo usam dados para aprender mais sobre o mundo seja no sentido informal das observações diárias ou, principalmente nos últimos séculos, no sentido formal de unidades quantificadas que podem ser manipuladas por potentes algoritmos. Os exemplos usados por Schönberger- Mayer e Cukier (2013) se referem aos conceitos de *censo* e *estatística*, respectivamente a coleta e análise de informações da sociedade. Esses conceitos são muito utilizados nos dias de hoje com finalidades diversas tanto para governos quanto para a sociedade.

Como aborda Burke (2003), nem mercantilizar a informação nem utilizá-la como forma de reprimir, dominar e expressar poder são características de cunho político, social e econômico recentes, e sim tendências vindas de longo prazo. Os governos desde as épocas mais remotas têm se interessado por informações sobre suas populações e riquezas para fins

militares e tributários. Há indícios na história de que 3000 anos a.C. já se faziam censos na Babilônia, China e Egito e até mesmo o 4º livro do Velho Testamento faz referência a uma instrução dada a Moisés, para que fizesse um levantamento dos homens de Israel que estivessem aptos para guerrear (Bíblia Sagrada: Livro dos Números: 1, 1-3: p.175). Usualmente, estas informações eram utilizadas para a taxaço de impostos ou para o alistamento militar. Outro registro da Bíblia informa que o Imperador César Augusto, ordenou que se fizesse o censo de todo o Império Romano (Bíblia: Lucas: 1, 1-3: p.1530) (SCHÖNBERGER-MAYER; CUKIER, 2013).

No passado, a tarefa de coletar e analisar informações só cabia a instituições mais poderosas, como a Igreja e o Estado. Burke (2003) considera a crescente centralização de poder nos Estados e igualmente a de documentos e livros em prédios construídos especificamente para o armazenamento e consulta pública desse material, como arquivos e bibliotecas, nos principais centros urbanos europeus. A burocratização dos Estados, as sucessivas medidas de controle das informações, os mapeamentos, os questionários e relatos de expedições, o surgimento da estatística são alguns fatores analisados por Burke (2003)⁸.

Ao longo de milênios, os governos tentaram manter o controle da população coletando informações; o censo é citado por Schönberger-Mayer e Cukier (2013) como um exemplo disso. A palavra "censo" é derivada da palavra "*censere*", que em Latim significa "taxar". Segundo os autores, conduzir um censo na época era caro e demorado, conforme exemplificado pelo livro *Domesday*, de 1086, um dos mais venerados tesouros britânicos. Essa obra foi uma contagem abrangente – e sem precedentes em seu tempo – dos ingleses, suas terras e propriedades. Schönberger-Mayer e Cukier (2013) lembram que funcionários se espalharam pelo interior coletando informações para inscrevê-las no livro. Os resultados desse censo foram publicados em 1086 no livro intitulado "*Domesday Book*" e serviram de base para o cálculo de impostos. O livro era chamado de "fim dos dias" porque o processo era como o Juízo Final bíblico, no qual todas as vidas eram contadas. O Rei Guilherme I, que encomendou o livro, não viveu para vê-lo completo, e mesmo depois de todo o tempo e dinheiro investidos, as informações não eram precisas, eram apenas aproximadas, já que os funcionários não podiam contar todos os dados com perfeição.

Como já visto, a coleta sistemática de informação por parte dos governos passou a adquirir importância crescente desde a segunda metade do século XVIII. Segundo Burke (2012), grande parte dessa informação provinha de levantamentos oficiais do mais variados

⁸ Informações mais aprofundadas no livro *Uma história social do conhecimento: de Gutenberg a Diderot*, de Peter Burke (2003).

tipos, inclusive os censos, sendo que uma parte desses dados era fornecida por agentes secretos “informantes” ou espiões⁹. “A espionagem é uma velha atividade, mas foi em nosso período que ela se profissionalizou, no sentido de se tornar uma ocupação em tempo integral” (BURKE, 2012, p.156).

Todavia, Burke (2012) ressalta que, desde a segunda metade do século XVIII, os governos tendem a destinar uma proporção maior de seus orçamentos de segurança para a coleta de informação sobre a própria população. Segundo ele, os franceses foram os pioneiros nesse aspecto, seguidos pelos prussianos, austríacos e russos.

Outra razão tradicional para os Estados colherem informações sobre os indivíduos é a vontade de controlar os movimentos, tanto de entrada de estrangeiros quanto a saída de seus cidadãos. Os governos dos países que atraíam um maior fluxo de imigrantes, como os Estados Unidos entre as décadas finais do século XIX e as iniciais do século XX, procuravam informações que ajudassem a selecionar quem poderia entrar (BURKE, 2012, p. 160).

Outro exemplo que ilustra esse item diz respeito ao aparecimento do conceito de estatística. A origem da palavra estatística está associada à palavra latina *status* (Estado). Há mais de 300 anos, um comerciante britânico chamado John Graunt teve uma ideia considerada inovadora para a época. Ele queria saber o tamanho da população de Londres na época da peste negra (1665)¹⁰. Segundo Schönberger-Mayer e Cukier (2013) em vez de contar todas as pessoas, Graunt inventou uma abordagem que lhe permitiu estimar a população, hoje chamada de *estatística*. Com uma abordagem simples, ele estabeleceu a ideia de que uma pessoa podia extrair conhecimento da população em geral a partir de uma pequena amostra. Essa prática tem sido continuada nos tempos modernos.

Segundo Burke (2012), o conhecimento relacionado à política social passou a ser cada vez mais necessário desde a segunda metade do século XIX, à medida que os Estados começaram a se preocupar mais com o bem-estar de seus cidadãos nas áreas de saúde, educação e lazer. A Suécia é pioneira na coleta de informações para finalidades oficiais. No

⁹ O método tradicional usado pelos governos para adquirir informação sobre países estrangeiros, fosse para objetivos governamentais ou para a movimentação das tropas, se dava por intermédio dos embaixadores, que por sua vez dependiam de uma rede de informantes. Mas foi apenas no século XX que os governos, na maioria, complementaram ou trocaram os métodos tradicionais especializados para os serviços secretos (BURKE, 2012, p. 156 - 157).

¹⁰A Grande Praga de Londres foi um surto de peste bubônica. Houve outros antes dela na Inglaterra e ao redor do mundo. O mais conhecido foi a pandemia chamada de Peste Negra, no século 14. Especialistas apontam que ela possa ter dizimado 1/3 da população europeia na época, aproximadamente 75 milhões de pessoas. A Grande Praga matou um número estimado de 100 mil pessoas, equivalente a 20% da população da capital britânica em 1665. Acredita-se que os focos da doença só tenham sido eliminados em 1666, com o Grande Incêndio, que destruiu grande parte das favelas e lugares mais pobres da cidade.

século XIX, colheram-se estatísticas sobre a agricultura, o comércio, a manufatura e as exportações, e em 1858 foi criado um Escritório Central de Estatística.

Sobre isso Burke diz:

Na segunda metade do século XIX, a industrialização de uma sociedade até então camponesa, ao lado da ampliação do direito do voto e do surgimento de organizações políticas em grande escala, incentivou o Estado a desempenhar um papel social mais ativo do que antes. Para preparar o terreno, o governo utilizou comissões de investigação ad hoc, cerca de 4 mil ao todo entre 1855 e 1974, para colher informação sobre os números de desempregados, de idosos e assim por diante (BURKE, 2012, p. 161).

Burke (2012) afirma que a partir dos relatórios dessas comissões, criaram-se pensões para os idosos, seguros-saúde e auxílio-desemprego. No entanto, as implementações dessas medidas exigiram uma quantidade ainda maior de informação.

Como abordado anteriormente, os censos sempre foram importantes fontes de conhecimento da população para os governos. Em 1880, os Estados Unidos realizaram um censo nacional de 50 milhões de pessoas que recolheu informações demográficas, incluindo idade, sexo, número de pessoas no domicílio, etnia, data de nascimento, estado civil, profissão, estado de saúde, alfabetização e lugar de origem. Toda essa informação foi registrada à mão, microfilmada e enviada para armazenamento em arquivos estaduais, bibliotecas e universidades. Demorou de sete a oito anos para tabular corretamente os dados do censo após a coleta inicial. Em 1890, o Census Bureau dos EUA racionalizou seus métodos de coleta de dados através da adoção de cartões perfurados legíveis por máquina que poderiam ser tabulados em um ano. Especialistas apontam o *Social Security Act*¹¹ — programa instituído pelo presidente dos EUA Franklin Roosevelt em 1937 — como o primeiro grande projeto de captação e catalogação de informações. Por meio de um sistema de cartões perfurados, idealizado e implantado pela IBM, o governo computou informações de 26 milhões de americanos. Já no censo mais recente dos EUA, realizado em 2010, o departamento usou uma gama de tecnologias emergentes para examinar a população, incluindo os sistemas de informação geográfica, mídia social, vídeos, sistemas inteligentes de reconhecimento de caracteres e softwares sofisticados de processamento de dados.

¹¹A segurança social é um conceito consagrado no artigo 22 da Declaração Universal dos Direitos Humanos que afirma que toda pessoa, como membro da sociedade, tem direito à segurança social e tem direito à realização, pelo esforço nacional, pela cooperação internacional e de acordo com a organização e recursos de cada Estado, dos direitos econômicos, sociais e culturais indispensáveis à sua dignidade e ao livre desenvolvimento de sua personalidade. Em termos simples, os signatários concordam que a sociedade em que vive a pessoa deve ajudá-los a desenvolver e tirar o máximo partido de todas as vantagens (cultura, trabalho, bem-estar social), que são oferecidos a eles no país. Informações retiradas do http://en.wikipedia.org/wiki/Social_security_system

Já no Brasil durante o período imperial, o governo sentiu necessidade de obter dados estatísticos para conhecer melhor o País¹². Por conta disso, em 1871, criou a Diretoria Geral de Estatística – DGE, para organizar as atividades estatísticas nacionais e realizar, no ano seguinte, o primeiro recenseamento feito no Brasil. Em 1872 foi realizado o primeiro censo nacional que recebeu o nome de Recenseamento da População do Império. Neste Censo, a região Nordeste ainda aparecia como a maior do país na segunda metade do século XIX. Dos oito estados mais populosos do Brasil, metade eram nordestinos como mostra o Quadro 1. Esses estados também estavam dentre os mais povoados, já que Pernambuco tinha área menor que São Paulo, mas com população maior e a Paraíba com quase a mesma população do Rio Grande do Sul e área bem menor também. O Brasil tinha 9.930.478 habitantes, sendo 5.123.869 homens e 4.806.609 mulher.

Quadro 1: Censo populacional brasileiro de 1872

	ESTADO	HABITANTES
1º	Minas Gerais	2 milhões
2º	Bahia	1,4 milhões
3º	Pernambuco	1,1 milhão
4º	Rio de Janeiro	1 milhão
5º	São Paulo	800 mil
6º	Ceará	700 mil
7º	Rio Grande do Sul	450 mil
8º	Paraíba	400 mil

O censo de 1872 também foi o único da história do Brasil que contabilizou os escravos, já em declínio. Além de questões básicas sobre sexo, idade, escolaridade, estado civil, religião e deficiência física, o questionário incluía a pergunta se a pessoa era livre ou escrava.

¹²Informações sobre a história dos censos retirados do site do IBGE. Disponível em: <http://memoria.ibge.gov.br/pt/sinteses-historicas/historicos-dos-censos>. Acesso em: 25 de setembro de 2014.

No entanto, desde 1750 já havia informações oficiais sobre a população no Brasil por ordem da Coroa Portuguesa. Antes da realização do primeiro censo nacional, foram realizados vários censos de caráter estadual e municipal. No Rio de Janeiro foram realizados censos nos anos de 1799, 1821, 1838, 1849, 1856 e 1870. Em São Paulo, nos anos de 1765, 1777, 1798 e 1836. Em 1831, o estado de Minas Gerais realizou censo contendo listas nominativas dos habitantes.

Hoje os recenseamentos são realizados a cada 10 anos pela Fundação IBGE (Instituto Brasileiro de Geografia e Estatística), órgão responsável pelas estatísticas (dados estatísticos) oficiais brasileiras como mostra o Quadro 2 e têm finalidades diversas em relação à época do Brasil Império, que eram realizadas para finalidades de cobranças de impostos e recrutamento militar. As informações coletadas sobre a população brasileira na atualidade definem as políticas públicas nos âmbitos nacional, estadual e municipal. Os resultados também ajudam a iniciativa privada a tomar decisões sobre investimentos.

Quadro 2: Censos Demográficos do Brasil

ano	população	crescimento no período
1872	9.930.478	
1890	14.333.915	44,3%
1900	17.438.434	21,6%
1920	30.635.605	75,6%
1940	41.236.315	34,6%
1950	51.944.397	25,9%
1960	70.191.370	35,1%
1970	93.139.037	32,6%
1980	119.002.706	27,7%
1991	146.825.475	23,3%
2000	169.799.170	15,6%
2010	190.732.694	12,3%

Fonte: IBGE (2015)

O último censo do Brasil aconteceu em 2010, registrando mais de 190 milhões de habitantes. O recenseamento inovou na coleta e nas perguntas feitas à população, com pesquisa totalmente informatizada: com computadores de mão, equipados com receptores

GPS e mapas digitalizados. A pesquisa contou com a participação de 230 mil agentes¹³ e aconteceu em 67,6 milhões de domicílios, distribuídos por todo país.

Assim, a partir desse resgate, é possível observar que a medição da realidade e o registro de dados prosperaram por conta de uma combinação de instrumentos e de uma mentalidade receptiva. Essa combinação é o solo fértil no qual a datificação moderna cresceu. Segundo Schönberger-Mayer e Cukier (2013), a datificação já existia muito antes da invenção dos computadores ou de instrumentos tecnológicos. O empenho em compreender a natureza por meio da quantificação definiu a ciência no século XIX, quando eruditos inventaram novos instrumentos e unidades para medir as correntes elétricas, a pressão do ar, a temperatura, a frequência do som. Essa foi uma era em que absolutamente tudo tinha de ser definido, demarcado e denotado. E hoje essas práticas se tornaram intencionais na busca constante por conhecimentos através de dados e informações que estão sendo geradas, principalmente nos ambientes digitais. Assim, observa-se que a busca dos conhecimentos por meio de dados e informações é algo longe se acabar, já que o homem tem tanta necessidade de saber sobre a si mesmo.

2.4.1 O conceito de datificação

A importância da coleta de dados para buscar conhecimentos tem passado de séculos para séculos. No entanto, hoje os dados que estão sendo coletados não têm o mesmo significado que antigamente. Os dados não são usados somente para usos primários, mas são reaproveitados para outras finalidades como observam Schönberger-Mayer e Cukier.

Eles não são mais considerados estáticos e banais, cuja utilidade terminava depois que o objetivo da coleta era alcançado. Em vez disso os dados se tornaram matéria-prima dos negócios, um recurso econômico vital, usado para criar uma nova forma de valor econômico. Na verdade, com a mentalidade certa, os dados, podem ser reutilizados para se tornarem fonte de inovação e novos serviços. Eles podem revelar segredos para aqueles com humildade, disposição e instrumentos para ouvir (SCHÖNBERGER- MAYER; CUKIER, 2013, p.4).

A essa nova ordem na contemporaneidade é de pensar os dados como possíveis de reutilizações e com o desenvolvimento de tecnologias cada vez mais sofisticadas para coletar, medir, armazenagem e analisar esses dados fez surgir a terminologia de datificação. A datificação, em inglês *datafication*, é considerado um fenômeno da explosão informacional que consiste em transformar informações em dados (SCHÖNBERGER-MAYER; CUKIER,

¹³ A autora dessa tese atuou como uma das agentes na cidade do Rio de Janeiro, fazendo recenseamento populacional do Bairro da Urca (Zona Sul).

2013). Datificação se refere à coleta de informações de boa parte que existe, seja no mundo digital ou analógico, transformando tudo em dados que possam ser quantificados de modo que possa ser tabulado e analisado. Esse processo permite usar as informações de novas maneiras, como análise previsível. Seria assim, um processo sistemático de extração e tabulação de dados dispersos e aparentemente irrelevantes, mas que sempre que a datificação for bem sucedida pode-se revelar muito valor agregado pelas informações subjacentes.

Assim, datificação se configura como:

À coleta de informações de tudo o que existe – inclusive informações que nunca foram pensadas como tal, como a localização de uma pessoa, as vibrações de um motor ou o estresse estrutural numa ponte – e à transformação disso em dados que possam ser quantificados (SCHÖNBERGER- MAYER; CUKIER, 2013, p.10).

Para os autores a datificação permite usar as informações de novas maneiras, “como na análise previsiva: detectar que um motor está prestes a quebrar com base no calor ou vibrações que ele gera. Como resultado, revelamos o valor latente e implícito das informações” (SCHÖNBERGER- MAYER; CUKIER, 2013, p.10).

Assim a ideia da datificação é uma das características de várias empresas de mídias sociais. “As plataformas de redes sociais não apenas nos oferecem uma maneira de encontrar e manter contato com amigos e colegas, mas usam elementos intangíveis do cotidiano e os transformam em dados que podem ser usados para outros fins” (SCHÖNBERGER- MAYER; CUKIER, 2013, p 63).

O Facebook datifica a rede de amigos, o Google tem datificado a busca e recuperação de informações. Twitter é um datificador de notícias e de informações em tempo real, o LinkedIn, por exemplo, tem datificado os contatos profissionais. Dessa forma cada uma das empresas de relacionamento busca aproveitar essa grande quantidade de dados não estruturados (*big data*) para armazenar, analisar e rentabilizar a informação em torno das suas atividades e negócios.

Cada informação é “datificada”, com o objetivo de monitorar, armazenar e começar a identificar padrões de uso e, portanto, ter recomendações de como melhorar ou monetizar os negócios. Por exemplo, os executivos de diversos segmentos da indústria têm tentado analisar e “datificar” vendas, aquisição de clientes, rentabilidade de produtos e custos da cadeia de suprimentos há anos.

A novidade é que, agora, esse processo tem velocidade rápida e ferramentas adequadas para armazenar, manipular e analisar determinadas informações. Campanhas políticas também têm usado dessas práticas, por meio de monitoramento de redes sociais, buscam a datificação das informações dos eleitores presentes nesses ambientes. Assim, como empresas que trabalham

com pesquisas a partir de análises de dados de mídias sociais, buscam por meio de perfis de consumidores e usuários a datificação de seus bios (biografia do perfil) (Figura 6).

Figura 6: Bio do Twitter



Fonte: Página do Twitter da autora (2015)

Nas “bios” é possível encontrar informações importantes sobre o usuário da rede, como quem é, o que faz, do que gosta. É isso que faz com que outros usuários decidam acompanhar o perfil ou não do usuário da rede.

Schönberg-Mayer e Cukier (2013) chamam a atenção para as diferenças entre os conceitos de datificação e digitalização. Para eles digitalização é apenas um processo de representação numérica por meio de codificação binária. O ato de digitalização, ou seja, a transformação de informações analógicas no formato que os computadores possam ler em si não datifica. Um exemplo usado por esses autores quanto a essa diferença é da digitalização de livros a partir do Google Books: a página digitalizada é transformada em uma imagem ou similar, mas isso não torna o texto em si um dado. É necessário que se façam outros processos (como o reconhecimento textual) para que as palavras em si e a coerência que elas mantêm possam ser reconhecidas como dados.

A Google¹⁴ digitalizou os textos: todas as páginas foram escaneadas e transformadas numa imagem de alta resolução armazenadas nos servidores da empresa. Cada página foi transformada numa cópia digital que podia facilmente ser consultada na Internet. Para recuperá-la, contudo, era preciso saber em que livro estava a informação ou ler muitas páginas até encontrá-la. Não era possível procurar determinadas palavras no texto, nem analisá-lo, porque o texto não fora dataficado. Tudo o que o Google tinha eram imagens que apenas os seres humanos podiam transformar em informações úteis – isto é, lendo. (SCHÖNBERGER- MAYER; CUKIER, 2013, p.57-58)

¹⁴ A Google é referida aqui no texto no contexto feminino porque se refere a empresa.

Para além dessa necessidade, os autores sublinham a importância da medição e do registro como facilitadores de criação de dados. Sem a invenção da escrita, como instrumento de catalogação, e a adoção dos algarismos, voltados ao cálculo, seria inviável qualquer processo de captura e aglutinação de dados.

A datificação permitiu a indexação e a busca nos textos, além de um fluxo infinito de análise textual. Estima-se que 130 milhões de livros tenham sido publicados desde a invenção da imprensa no século XV. Em 2012, o Google já havia escaneado mais de 20 milhões de títulos, mais de 15% de toda a herança escrita do mundo.

Schönberger-Mayer e Cukier (2013) destacam que a transformação de palavras em dados gera inúmeros usos. Os dados podem ser usados pelos seres humanos para a leitura e pelas máquinas para a análise. Para esses autores (SCHÖNBERGER-MAYER; CUKIER, 2013, p.57) “o surgimento dos computadores gerou aparelhos de medição e armazenagem digitais que tornaram a datificação muito mais eficiente. Eles também permitiram que a análise matemática dos dados revelasse seu valor oculto”.

Nas redes sociais como Facebook e Twitter é possível fazer coletas e análises de textos sobre opiniões e sentimentos. Assim essas empresas se tornam importantes ambientes de comunicação que agrupam diversas informações, entre elas opiniões e sentimentos expressos por seus usuários em simples conversas, que têm sido usadas para diversas finalidades tanto comerciais quanto para pesquisas científicas. Dentre as redes sociais, o Twitter tem sido a rede predileta para esse tipo de pesquisa (TSYTASARAU, 2010), porque permite aos usuários enviar e receber atualizações pessoais de outros contatos. Porém, o grande problema visto neste tipo de pesquisa é que a maioria dos autores que publica na área de análise de sentimentos cria a sua própria base de dados e muitas vezes não a disponibiliza. Segundo Filho *et al* (2015), a técnica mais utilizada e com os melhores resultados na análise de sentimentos tem sido a aplicação de algoritmos de aprendizado de máquina¹⁵.

Nessa discussão, coloca-se em evidência o poder que as provedoras de sites de relacionamento têm sobre os dados pessoais dos seus usuários que são usados para diversas

¹⁵ Aprendizado de máquina é um sub-campo da inteligência artificial dedicado ao desenvolvimento de algoritmos e técnicas que permitam ao computador aprender, isto é, que permitam ao computador aperfeiçoar seu desempenho em alguma tarefa. Enquanto que na inteligência artificial existem dois tipos de raciocínio - o indutivo, que extrai regras e padrões de grandes conjuntos de dados, e o dedutivo - o aprendizado de máquina só se preocupa com o indutivo. Algumas partes da aprendizagem automática estão intimamente ligadas à mineração de dados e estatísticas. Sua pesquisa foca nas propriedades dos métodos estatísticos, assim como sua complexidade computacional. Sua aplicação prática inclui o processamento de linguagem natural, motores de busca, diagnósticos médicos, bioinformática, reconhecimento de fala, reconhecimento de escrita, visão computacional e locomoção de robôs.

finalidades, inclusive comerciais. Dessa forma, entende-se que o processo de datificação tem sido os grandes bancos de dados das empresas de mídia social que formam a base de novos empreendimentos e que vão muito além do superficial compartilhamento de fotos, atualizações de status e botões de “curtir”. São importantes ambientes de informações que podem ser transformados em dados com valores estratégicos, que podem ser aplicados para diversas finalidades, mas que também provocam questionamentos em relação à ética e privacidade dos usuários, o que será melhor explorado no capítulo 4 desta tese.

No capítulo a seguir contextualiza-se a temática *big data*, que é o objeto de investigação deste trabalho. Procura-se por meio de diferentes discussões teóricas e práticas apresentar o conceito e suas características a partir dos chamados 3 Vs (Volume, Variedade e Velocidade). Além disso, são citados alguns exemplos de usos do *big data* por governos e empresas, principalmente para finalidades de vigilância e controle de dados e informações dos indivíduos.

3 *BIG DATA*: DIFERENTES DIMENSÕES CONCEITUAIS E PRÁTICAS

Traçadas as noções de informação, datificação no contexto da sociedade informacional, o presente capítulo tem como finalidade discutir o conceito e aplicabilidade do *big data*, considerando discussões teóricas e práticas que cercam a temática. Aqui optou-se por tratar inicialmente o tema sob um aspecto amplo, por considerar e identificar diante das pesquisas e levantamento da literatura, que pouco ainda têm se discutido sobre *big data* em âmbito acadêmico principalmente nas áreas de informação e comunicação.

As tecnologias digitais têm ocupado um grande espaço dentro da vida cultural, social, política e econômica. Tentar compreender, contextualizar e situar narrativas, práticas de produção e análise de dados de rede é um grande desafio para a sociedade e, sobretudo, para os cientistas sociais. O professor de Ciências Políticas da Universidade de Stanford, Justin Grimmer, no artigo publicado no jornal *Political Science and Politics*¹⁷, reflete sobre a questão do “Big Data” como um novo campo de pesquisa e possibilidades para os cientistas sociais. Aqui destacam-se algumas dessas pesquisas em âmbito nacional e internacional. Em seguida contextualiza-se a temática, conceituando e caracterizando, bem como dando destaque para algumas áreas que já se utilizam na prática do *big data*. Não se tem a pretensão neste estudo demonstrar exaustivamente como o *big data* funciona do ponto de vista tecnológico e instrumental. Esse não é o foco da presente investigação. Assim, busca-se um direcionamento mais estratégico do *big data*, a fim de explorar suas possibilidades, limites e riscos a partir das análises empreendidas de dados coletados de diferentes fontes e formatos.

No que se refere a pesquisa bibliográfica para contextualizar *big data*, usa-se principalmente as contribuições dos autores Schönberger-Mayer e Cukier (2013), que são os pioneiros e referências no mundo na discussão dessa temática.

3.1 *BIG DATA* NAS CIÊNCIAS SOCIAIS

O mundo está cada vez mais complexo e os velhos modelos de abstração não conseguem dar conta da crescente sofisticação e interligação dos sistemas digitais. Essas são considerações feitas por Radfahrer (2012).

De uma rede de comunicação global a internet evoluiu para uma espécie de memória coletiva, a que boa parte dos processos é delegada. Smartphones, computação em nuvem e internet das coisas, cada vez mais familiares, mensuram cada transação e por meio da compilação de dados definem quem é você, o que faz e onde passa cada instante (RADFAHRER, 2012).

¹⁷ Disponível em: http://stanford.edu/~jgrimmer/bd_2.pdf. Acesso em: 20 de Fevereiro de 2015.

O trabalho com *big data* vem ganhando cada vez mais espaço em campos como comunicação e tecnologia, impulsionado principalmente pela produção de dados e informações em larga escala, pela sua disponibilidade na internet e pelo aumento das ferramentas para extração e análise desses dados. Raquel Recuero (2012) observa que há uma gama de pesquisadores de todo o mundo interessado no que parece ser a grande novidade em termos de dados, proporcionada principalmente pelo advento das ferramentas de comunicação mediada por computador.

O *big data* para alguns pesquisadores pode transformar as ciências humanas. “O século 21 ficará conhecido como século em que a humanidade quis entender seu próprio comportamento”. A reflexão é do historiador canadense Chad Gaffield, que ministrou palestra *Understanding People In The Digital Age* na Universidade de São Paulo (USP) em 2013. Para o especialista e presidente do *Social Science and Humanity Research Council* (SSHRC), organização dedicada à pesquisa das ciências sociais no Canadá, o mundo vive o período de maior mudança desde a invenção da imprensa. “Percebemos que o foco agora é ampliar o conhecimento sobre pessoas”, explica Gaffield.

Chad Gaffield dedicou-se nos últimos anos aos estudos de dados, mais especificamente do *big data*. “Nos negócios o foco começou a mudar do produto para o consumidor. Hoje, empresas querem saber o que as pessoas querem, os produtos são ‘*user-oriented*’ e os serviços buscam entender como o usuário pensa”, declara ao citar previsões sobre como o conhecimento sobre o homem pode contribuir para vida em sociedade.

No trabalho, as empresas terão o dever de compreender seus empregados. Assim como os políticos, cada vez mais, terão que entender quem é o cidadão. Na educação, por 200 anos, não se preocuparam em entender como as pessoas aprendem. Sabemos hoje que transmitir conhecimento unilateralmente nem sempre é o melhor jeito de aprender. Na saúde, ouvir o médico era a principal fonte de informação, hoje a saúde é sobre o paciente e não mais sobre a doença (GAFFIELD, 2013)¹⁸

Todo esse interesse, segundo Gaffield, é resultado direto de uma época em que a humanidade se encontra cada vez mais conectada, o que nos permite ter informações sobre outros como nunca antes. “O aspecto mais distinto sobre o *big data* deve ser esse: o fato desses grandes conjuntos de dados serem sobre a humanidade, sobre pessoas”. Para o pesquisador é a partir da análise das relações entre pessoas que poderemos entender como as mudanças acontecem na sociedade. Da mesma forma cientistas como Stephen Wolfram e

¹⁸ Palestra realizada na USP em 2013. O evento reuniu, na Faculdade de Filosofia, Letras e Ciências Humanas (FFLCH) da USP, professores, cientistas sociais e interessados no conhecimento que a análise de dados por sistemas digitais pode gerar.

laboratórios como o MIT *Connection Science* propõem uma integração da teoria de redes, computação e ciências comportamentais para propor novos modelos de fluxo de informação e influência em redes sociais, sejam elas digitais ou não.

Albert-Laszló Barabasi (2002)¹⁹, considerado uma das maiores referências mundiais no estudo de redes complexas, foi o primeiro a utilizar dados digitais para tentar compreender o comportamento das pessoas na Internet. Ao encontrar padrões em rastros de navegação, concluiu que redes complexas possuem propriedades e que essas propriedades podem ser aplicadas no conhecimento e gestão de outros sistemas complexos como um organismo, uma célula ou a própria sociedade.

Sobre isso Costa e Cavalcanti abordam que nos últimos quatro séculos, o método científico foi usado para construir o conhecimento.

Olhando para fora de nós fomos capazes de identificar quase todas as partes que compõem o universo. Descobrimos novas espécies, novas partículas e elementos químicos e novos planetas e galáxias. Olhando para dentro, descobrimos todos os órgãos e seus componentes, todas as substâncias que compõem o corpo humano, todas as partes de uma célula e sequenciamos todo o genoma humano. Esse avanço do conhecimento foi resolvendo alguns problemas e trazendo novos desafios, novas zonas de sombra e ignorância. Se em algum momento achávamos que íamos descobrir a explicação definitiva do universo, a vida rapidamente se encarregava de nos colocar novos problemas, bem mais complexos (COSTA; CAVALCANTI, 2014).

Para esses autores, a ampliação do conhecimento científico nos leva ao desafio de que conhecemos as mínimas partes de todo sistema que estudamos, mas temos cada vez mais problemas sem solução. Dessa forma, eles entendem que a Ciência das Redes traz novo oxigênio ao campo das ciências, complementando o método científico reducionista que chegou ao seu limite. “Conhecer as partes nos possibilitou muito conhecimento e inovação, mas garantiu a solução apenas daqueles problemas que têm a sua gênese no entendimento das partes do sistema, e não na relação entre estas partes” (COSTA; CAVALCANTI, 2014).

Não são poucas as pesquisas em âmbitos acadêmicos que têm se dedicado a entender e analisar o grande volume de dados que a humanidade vem gerando nos últimos anos e que tem se mostrado como um grande desafio não somente para o mercado, mas também para a ciência. Alguns estudos, como se observou, valem-se da terminologia *big data* e outros preferem usar dados digitais.

Recuero (2012) por sua vez cita os estudos do Axel Bruns, na Austrália, que tem usado *big data* para entender o comportamento das pessoas nas mídias sociais em momentos de crise (por exemplo, durante um desastre) e tentando aprender com isso como criar

¹⁹ Ver A Nova Ciência das Redes (2002).

estratégias para lidar com essas crises. Ele também tem estudado como as mídias sociais (o que as pessoas falam na Internet, basicamente) têm a ver com as audiências feitas nas redes, tentando entender um pouco melhor o comportamento desses públicos

Outros estudos usando *big data* em análise de mídias podem ser apontados a partir das experiências internacionais, como as citadas por Ronaldo Cristiano Prati (2014). O sistema Lydia desenvolvido por Lloyd, Kechagias e Skiena (2005) é um sistema de propósito múltiplo que foi usado para detectar padrões espaciais e temporais de entidades presentes nas notícias, a comparação de ocorrência dessas entidades entre mídias tradicionais e blogs, o sentimento associado a essas entidades e a sua correlação com o mercado financeiro para a previsão de preço de ações. Esse sistema também foi utilizado nas análises das campanhas presidenciais americanas de 2008 e 2012. Ali *et al.* (2011) desenvolveram o sistema NOAM para coleta e análise automática de notícias. Segundo Prati (2014), estudo reportado em Flaounas *et al.* (2013) descreve uma ampla pesquisa com 2,5 milhões de notícias coletadas a partir de 498 diferentes provedores de 98 países e que publicaram textos na língua inglesa por um período de 10 meses. As notícias foram automaticamente classificadas utilizando técnicas de aprendizado de máquina em 15 categorias diferentes. O objetivo era encontrar padrões e similaridades no estilo de escrita entre diferentes provedores de notícias, diferenças de gênero no conteúdo das notícias e o seu relacionamento e destes com a popularidade dos artigos.

Além de notícias atuais, o *big data* também vem sendo aplicado para analisar arquivos históricos de jornais. Um exemplo é o projeto “*An Epidemiology of Information*” que tem como proposta investigar padrões de disseminação de informação a respeito da epidemia de gripe espanhola nos Estados Unidos. De acordo com Prati (2014) os pesquisadores Ewing, Gad e Ramakrishnan (2013) usaram técnicas de inteligência artificial para estudar o padrão de cobertura jornalística a respeito dos locais em que focos de gripe ganhavam a atenção da mídia da época. Posteriormente, Ewing *et al.* (2014) estudaram o tom das notícias, classificando-as como alarmistas, de aviso, explanatórias e tranquilizadoras.

Outro exemplo de pioneirismo nesse campo, envolvendo o processamento e análise de grandes volumes de imagens na internet, é o Laboratório *Software Studies*, da Universidade da Califórnia, criado em 2007 pelo professor e pesquisador Lev Manovich. O pesquisador e sua equipe desenvolvem ferramentas que possibilitam a visualização e, conseqüentemente, a análise de grandes volumes de imagens. O destaque de suas pesquisas é a criação do ImagePlot, um plugin desenvolvido para um software chamado ImageJ, que permite a organização de milhares de imagens em um único gráfico (ou *plot*), obedecendo a alguns parâmetros pré-determinados, dentre eles cor, brilho, saturação e tempo. As pesquisas do

Software Studies não possuem tema único, passeando por entre o campo artístico (comparação entre pinturas de Mondrian e Rocko, por exemplo) e atingindo análises de dados on-line (como o projeto *Phototrails*). Os parâmetros de cor, brilho e saturação que compõem essas visualizações dão margem às análises que vão desde poder revelar padrões comportamentais de uma sociedade até observar a variância da evolução de uma veia artística. Entre os projetos desenvolvidos pelo laboratório está o *Phototrails*, idealizado por Nadav Hochman, Lev Manovich e Jay Chow. O objetivo proposto com esse estudo foi explorar padrões visuais, dinâmicas e estruturas em escala planetária sobre o compartilhamento de fotos geradas por usuários, usando uma amostra de cerca de 2,3 milhões de fotos do *Instagram*, vindas de 13 cidades ao redor do planeta. As teorias formuladas a partir das análises das visualizações mostram como as mudanças temporais, de localidade e das características visuais podem revelar comportamentos sociais, culturais e políticos acerca da atividade das pessoas pelo mundo.

No Brasil existem já experiências em estudos com o *big data* que valem referências. O professor e pesquisador Fábio Malini, da Universidade Federal do Espírito Santo (UFES), tem trabalhado sobre as organizações e manifestações políticas e de ativismo nesses espaços de fluxos comunicativos. Desde 2012, o Laboratório de Estudos sobre Imagem e Cibercultura (Labic) vem recolhendo conteúdo público na Internet focado em manifestações políticas. O foco dos estudos de Malini está na cartografia política a respeito dos protestos brasileiros a partir de 2013. A pesquisa busca extrair dados de mídias sociais, especialmente do Twitter, de menções relacionadas com figuras políticas de destaque assim como *hashtags* utilizadas durante os protestos. A primeira análise de dados foi feita em 170 mil tweets, que representavam diversos perfis de mídia social no Brasil, variando de meios de comunicação tradicionais aos perfis dos movimentos sociais. Nesse estudo conseguiu-se fazer uma visualização de dados envolvendo determinados políticos e se descobriu dois padrões de comportamentos: os que são contra (negativo) e os que são favoráveis ao político (positivo), além desses um grupo que se colocava como neutro.

Da mesma forma, destacam-se as pesquisas do professor e pesquisador Walter Lima Júnior, do Observatório de Mídias Sociais Conectadas da Universidade Metodista de São Paulo. As suas pesquisas são focadas em redes sociais e campanhas eleitorais presidenciais. As primeiras pesquisas foram feitas sobre as eleições brasileiras de 2010 e 2014. Segundo Lima Júnior (2014) com o objetivo de compreender melhor as circunstâncias comunicacionais das campanhas on-line, surgiu o projeto de pesquisa aplicada Observatório de Mídias Sociais Conectadas, denominada de Neofluxo, elaborado pelo Grupo de Pesquisa Tecnologia,

Comunicação e Ciência Cognitiva (Teccog) e aprovado em junho de 2010 pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Por meio de *softwares* produzidos e/ou combinados, o projeto se dedicou aos seguintes propósitos interdependentes, verificando: (1) se mensagens dos programas televisivos dos principais candidatos à presidência eram utilizadas como parâmetro para processos comunicativos sustentados pelo *Twitter*, avaliando também as estratégias de tais candidatos na plataforma específica, (2) se havia ação de *Astroturfings* (usuários especializados que se passam por usuários comuns, por exemplo) vinculados partidos ou candidatos particulares. A partir da elaboração de critérios valor-notícia para seleção palavras-chave foi possível rastrear, coletar e armazenar participações de usuários cadastrados no *Twitter*, formando-se um banco de dados com mais de 20 milhões de mensagens. A pesquisa centrava suas análises a partir das redes sociais (Facebook, Flickr, Twitter, Youtube) e site oficial da campanha dos candidatos. Além disso, a pesquisa buscou mapear como aconteciam os fluxos informativos entre os canais oficiais dos candidatos e o *Twitter* (ponto de partida, caminho percorrido e redistribuição da mensagem), medindo os níveis de relação direta entre as mensagens emitidas pelos canais oficiais dos candidatos à Presidência e o conteúdo divulgado por outros usuários no *Twitter*. E identificava a presença e a influência de *astroturfings* (ação artificial de indivíduos que se passam por usuários comuns, disseminando boatos, mentiras ou informações vantajosas de determinados candidatos a outros usuários se passando por eleitores que não possuem relação direta com o partido ou postulante ao cargo de presidente) no processo de troca e disseminação de mensagens²⁰.

Na mesma iniciativa de pesquisas na Universidade Federal do Rio de Janeiro (UFRJ), foi criado em 2012 o Instituto *Big Data* Brasil, que trabalha com pesquisadores da COPPE/UFRJ em conjunto com outras Universidades, desenvolvendo projetos e pesquisas que têm como foco o uso dos grandes volumes de dados e informações extraídos dos ambientes digitais. Segundo Luciana Sodr  Costa²¹, o instituto desenvolve pesquisas inspiradas pela *Network Science*, a ci ncia que estuda as propriedades das redes cient ficas. Entre os projetos desenvolvidos est o os que s o mantidos em parcerias com a Funda o Get lio Vargas (FGV) e com a prefeitura do Rio. Al m disso, o Instituto oferece cursos livres, profissionalizantes e MBA para diversas empresas p blicas ou privadas. Segundo Sodr  (2015), o *Data Driven Inovation* demanda habilidades em mapeamento de sistemas

²⁰ Maiores informa es sobre essa pesquisa pode ser encontrada no livro Projeto Neofluxo- Atua o eleitoral do Astroturfing no fluxo informativo na web. Dispon vel em: <http://pt.slideshare.net/Teccog/projeto-neofluxo-atuao-eleitoral-do-astroturfing-no-fluxo-informativo-na-web>

²¹ Luciana Sodr  Costa declarou isso em entrevista para a autora desse trabalho.

complexos, gestão estratégica de dados, *design thinking* (fazer a pergunta certa, reduzir os dados e contar a história), visualização de dados, tomada de decisão e colocação do projeto em ação.

A partir das perspectivas apresentadas, conhecer e controlar os fenômenos complexos são as promessas do *big data* que têm mobilizado as pessoas em torno do tema. Entusiastas e críticos despendem grandes esforços para entender as possibilidades e consequências dessa nova plataforma de geração do conhecimento (COSTA; CAVALCANTE, 2013).

3.2 BIG DATA E SUAS PARTICULARIDADES

O termo *big data* tem recebido diversas contribuições para a sua conceituação na atualidade. Parte delas trata o termo como uma referência de mercado ou mesmo de “moda”, outros já consideram um conceito mal definido e pouco compreendido (TAURION, 2014), mas que merece ser estudado. Por ser ainda pouco explorado no meio científico, optou-se nesta tese usar o termo *big data* como popularmente vem sendo referenciado em Congressos e debates acadêmicos nacionais e internacionais, a fim de contribuir com as listas de discussões em torno da temática.

Nesse cenário, ainda em formulação, é importante destacar que o *big data* remete a própria possibilidade de datificação da sociedade, que leva a geração de mais e mais dados (TAURION, 2015). Dessa forma, a ideia básica é a percepção que praticamente tudo o que a humanidade faz no dia a dia vai gerar um “rastro digital” que poderá ser eventualmente analisado. Como observa Bruno (2012, p. 687) “comunicar é deixar rastros”. Assim qualquer ação digital, principalmente em espaços da internet é passível de ser coletada e analisada.

Por outro lado, Schönberger- Mayer e Cukier (2013, p.54) observam que apesar de haver uma crença implícita entre os tecnólogos de que *big data* remonta à revolução do silício, em essência, o avanço rumo ao *big data* é uma continuação da antiga busca da humanidade por medir, registrar e analisar o mundo. Para esses autores os sistemas modernos de tecnologia da informação (TI) possibilitam *big data*, mas não são os únicos responsáveis pelo seu surgimento. “A revolução da TI é evidente, mas a ênfase estava mais no *T*, na tecnologia. É hora de voltarmos o olhar para o *I*, a informação” (SCHÖNBERGER-MAYER; CUKIER, 2013, p.54). Nesse sentido, os autores abordam que a fim de captar informações quantificáveis, de datificar, é preciso saber como medir e registrar essa medição, o que exige um conjunto certo de instrumentos, além do desejo de quantificar e registrar. Como apontam eles, ambos são pré-requisitos da datificação, e desenvolvemos os fundamentos dessa prática séculos antes do nascimento da era digital, como abordado no capítulo anterior desta tese.

Ciências como a astronomia e a genômica, que vivenciaram uma explosão informacional nos anos 2000, cunharam o termo *big data* (SCHÖNBERGER-MAYER; CUKIER, 2013). Este termo, relacionado aos grandes volumes de dados, foi primeiramente citado no relatório "*Data, data, everywhere: a special report on managing information*", do periódico britânico *The Economist* (CUKIER, 2010). Entretanto, durante estes anos, o termo foi sendo utilizado e relacionado a *datawarehouses*²² ou soluções de *business intelligence*²³ (BI), com *data sets* de *terabytes* de dados. No entanto, *big data* representa muito mais que isto. Para Taurion (2014) tratar analiticamente o grande volume de dados pode gerar grandes benefícios para a sociedade e principalmente para as empresas que usarem essas análises. Na visão desse autor "o *big data* torna-se útil pela abrangência de dados que podem ser manuseados" (TAURION, 2014, p.122). Atualmente, o conceito está migrando para todos os campos do conhecimento humano (SCHÖNBERGER-MAYER; CUKIER, 2013).

É importante destacar que, com o grande universo originado ao redor do assunto nesses últimos anos, hoje o *big data* pode ser considerado uma região de conhecimento multidisciplinar que abrange diversas áreas produtivas e pesquisas científicas. É um campo focado em gerar conhecimento e inteligência a partir da extração de informações das análises de grandes quantidades de dados complexos. A despeito disso, Diebold (2012) destaca que o termo *big data* tende a evoluir de apenas um fenômeno momentâneo para uma nova disciplina, pois a necessidade de lidar com grandes volumes de dados e, principalmente, extrair informação destes dados vêm tornando-se termos-chave para a maioria das ciências.

Outros autores (Pospiec; Feldens, 2012) definem *big data* como um desafio em que há um equilíbrio desfavorável entre dados disponíveis e conceitos e tecnologias de informação atuais. Conseqüentemente, não há suporte para essa grande quantidade de informação, que segundo Callebaut (2012) está crescendo de tamanho, bem como está aumentando a rapidez desse crescimento.

O termo também é designado como uma nova grandeza informacional que procede tanto do aumento da capacidade de estocagem como da emergência de um novo tipo de saber que tais volumes de dados gerariam (BRUNO, 2014). A suposição é a de que quantidades massivas de dados, submetidas a procedimentos algorítmicos apropriados, revelariam regras inscritas em suas correlações, permitindo projetar padrões de comportamento e intervir no

²² Data warehouse é um depósito de dados digitais que serve para armazenar informações detalhadas relativamente a uma empresa, criando e organizando relatórios através de históricos que são depois usados pela empresa para ajudar a tomar decisões importantes com base nos fatos apresentados. Disponível em: <http://www.significados.com.br/data-warehouse/>>. Acesso em: 12 de dezembro de 2014.

²³ É um conjunto de técnicas e ferramentas para auxiliar na transformação de dados brutos em informações significativas e uteis a fim de analisar o negócio.

curso das ações de inúmeros indivíduos de modo a conjurá-las, quando indesejadas, ou incitá-las, se forem bem-vindas. Assim, trata-se de um modelo polivalente. Podem-se extrair, por exemplo, padrões de intenções de voto, de consumo ou de atentados terroristas, contanto que se trabalhe com imensas e heterogêneas bases de dados. Segundo Manovich (2013) grandes volumes de dados têm variado de algumas dezenas de *terabytes* para muitos *petabytes* de dados em um único conjunto de dados.

Assim, em linhas gerais, a partir das considerações dos autores que abordam a temática, *big data* é um termo guarda-chuva, que encobre diferentes ferramentas, metodologias, práticas e técnicas para solucionar desafios de analisar grandes quantidades de dados e que pode ser aplicado e usado em qualquer setor.

Apesar de ser referenciado como um tema emergente, o *big data* ocupa discussões desde o fim da década passada, mas nos últimos anos foi que o conceito extrapolou de vez os limites da academia e dos setores de tecnologia da informação (TI), ganhando notoriedade e destaque até mesmo na imprensa nacional, internacional, bem como em congressos.

Em 1997, durante conferência do Instituto de Engenheiros Eletricistas e Eletrônicos (IEEE) sobre virtualização, dois pesquisadores da NASA (*National Aeronautics and Space Administration* – Administração Nacional da Aeronáutica e do Espaço) publicaram o artigo *Application-controlled demand paging for out-of-core visualization* (COX, 2004) em que anteviam os novos desafios na área de computação ao ter que lidar com grandes volumes de dados. Pelo que se tem conhecimento, foi nesse artigo que o termo *big data* foi usado pela primeira vez. E a partir de 2008, as revistas especializadas vêm chamando a atenção para a “Era do *big data*”. A revista *Wired* em junho de 2008 abriu sua seção especial sobre “A Era Petabyte”, afirmando: “Nossa capacidade de captar, armazenar, e compreender grandes quantidades de dados está mudando a ciência, medicina, negócios e tecnologia. Como a nossa coleção de fatos e números cresce, assim será a oportunidade de encontrar respostas para as questões fundamentais”. Em fevereiro de 2010, *The Economist* começou seu relatório especial *Dados, Dados Everywhere* com a frase “a revolução industrial de dados” (cunhado pelo cientista da computação Joe Hellerstein). Na observação de Manovich (2013) o efeito está sendo sentido em todos os lugares, de empresa para a ciência, por parte do governo para as artes.

No *World Economic Forum* de 2012, foi citado o *big data* como uma nova classe de ativo econômico comparado ao petróleo. Ginni Rometty, CEO da IBM ressaltou que, da mesma forma que o petróleo foi o recurso natural que potencializou a última revolução industrial, dados serão os recursos que farão a diferença em termos de competitividade,

inovação e até mesmo sobrevivência nesta revolução industrial que estamos vivendo – a era da informação (LEONARDI, 2013). Ao fim, o fórum publicou o estudo "*Big data, grande impacto: novas possibilidades para o desenvolvimento internacional*", que mostra como o *big data* pode ser uma arma contra os problemas sócio-econômicos²⁴.

No Brasil, revistas (Veja, 2013; Ciência Hoje, 2013; Exame, 2012) publicaram reportagens apresentando o conceito de *big data* como a “revolução dos dados” ou como um “fenômeno *big data*”, ressaltando suas aplicações, principalmente nas áreas comercial e empresarial. Jornais nacionais (O Globo) também exploraram o tema em 2012. As reportagens buscaram contextualizar o tema com o objetivo de esclarecer os leitores sobre o que vem a ser *big data*.

O valor mais acessível de máquinas computacionais (processamento e memória) e dos dispositivos de captura e armazenagem de dados (sensores, câmeras fotográficas e de vídeo, celulares, pen-drives, flash memory, discos rígidos externos etc.) criaram inimaginável quantidade de dados que estão sendo disponibilizados na Web, proporcionando a formação da “Era do Big Data” (LIMA JUNIOR, 2011).

No entanto, é importante observar que a “era do *big data*”²⁵ só se materializou com a confluência de alguns fatores como a queda do custo de armazenar dados. Antigamente, guardar e acessar com eficiência um volume relativamente pequeno de informações exigia um investimento pesado em hardware, com o investimento financeiro considerável, além de a instalação demorar muito tempo. Nos anos 90, estocar 1 gigabyte custava 1000 dólares para as empresas. Hoje, custa 6 centavos, diz Frederico Grosso, diretor da unidade de softwares para análise de dados da HP, na América Latina (FEIJÓ, 2013). Assim, o *big data* é uma evolução natural da computação em nuvem, em que o poder de arquivamento e processamento das máquinas migrou para a internet. Os custos são rateados com milhares de usuários (FEIJÓ, 2013). “Devido ao modelo econômico adotado pela globalização, o termo *just in time*, totalmente dependente da necessidade de uma expansão virtual, se tornou a palavra de ordem das negociações e forçou a ampliação de estrutura para armazenamento de dados”²⁶. Um exemplo é o Hadoop, *software* de código aberto surgido dentro da *Google* que permite tratar grandes volumes de dados em tempo real²⁷.

²⁴Artigo: Big data se abre para o mundo. Disponível em: < <http://oglobo.globo.com/tecnologia/de-moneyball-davos-big-data-se-abre-para-mundo-4460918>>. Acesso em: 14 de Dezembro de 2014.

²⁵Veja. ano 46. n.20, p. 70-81, 15 de maio de 2013. Edição 2321

²⁶ Citação retirada da página da infoescola da autora Anna Adami. Disponível em: <<http://www.infoescola.com/informatica/big-data/>>. Acesso em: 15 de junho de 2013.

²⁷ Leia mais sobre esse assunto em <http://oglobo.globo.com/tecnologia/de-moneyball-davos-big-data-se-abre-para-mundo-4460918#ixzz2ISND5bXF>.

Já pensando o *big data* do ponto de vista social, para Recuero²⁸ (2013), uma das pesquisadoras da área de Internet, o termo se refere à imensidão de dados sobre indivíduos e grupos que hoje são gerados na Internet por esses atores. Recuero (2013) diz que,

[...]coletar dados de pessoas sempre foi um problema para a pesquisa. Pessoas mentem, inventam, são subjetivas e, o "pior", esses dados ainda eram "traduzidos" pelo pesquisador. Análise de redes sociais, por exemplo, era feita como no tempo do Moreno: perguntando às pessoas. Eis que com a Internet, centenas de milhares de pessoas passaram a criar e espalhar "rastros digitais". Passaram a deixar por aí registros de suas falas, seus gostos, seus pensamentos, seus amigos e etc. Centenas de milhares de dados passaram a ser gerados e publicados por todos nós. E o uso desses dados é o que "*big data*" se refere. É a primeira vez na História que cientistas têm acesso a dados brutos humanos "não subjetivos" e que essas análises podem ser feitas de forma quantitativa e em grande escala. E mais do que isso, de dados que são gerados em outras plataformas especialmente para isso (como dados de consumo, dados de comportamento e etc.) (RECUERO, 2013)²⁹

Para essa autora a importância do *big data* está na capacidade de se extrair informações dos mais variados lugares, principalmente dos "rastros digitais" deixados pelos indivíduos em suas ações diárias. E essa tem sido uma das possibilidades oferecidas pelas empresas de mídias sociais (Facebook e Twitter) tanto para finalidades mercadológicas quanto científicas, no que se refere aos estudos voltados para essa área do conhecimento.

3.2.1 Conceitos e características

Como se viu anteriormente, *big data* é um conceito relativamente novo, portanto sua definição sofre constantes aperfeiçoamentos, pois a consolidação do termo se dá através da percepção no tempo do seu valor para qualquer tipo de uso. A ideia central ao conceito é a tomada de decisão em tempo real sobre uma corrente contínua de dados (*streaming computing*), provenientes de diversas fontes.

Mesmo fora da literatura o conceito não segue uma única linha de pensamento. Dentre os especialistas entrevistados para este trabalho, ao se tirar um mínimo comum vimos que o *big data*, na sua raiz, fala em tratar um grande volume de dados com grande velocidade e a partir de uma variedade de informações (TAURION, 2014; BROWN, 2013). Outras definições para o termo são apresentadas na Tabela 1, porém sem rigidez conceitual.

²⁸ RECUERO, Raquel. *big data*: apontamentos e limitações. Postagem em 14.05.2013. Disponível em: <<http://www.raquelrecuero.com/arquivos/2013/05/big-data-apontamentos-e-limitacoes.html>>. Acesso em: 15 de junho de 2013.

²⁹ Dispo nível em: <http://www.raquelrecuero.com/>. Acesso em: junho 2014.

Tabela 1: Definições de *big data*

Autor	Definição
Cavalcanti (2015)	As pessoas, equivocadamente, associam Big Data apenas à grande quantidade de dados digitais hoje disponíveis. Na verdade, a maior riqueza do Big Data vem da variedade (diversidade) de dados. São dados transacionais, das redes sociais, gps... e dos assuntos os mais variados. Costumamos dizer que Big Data está associado a 4 "V"s: volume (quantidade), variedade (diversidade), veracidade (são dados coletados on line, na hora que estamos fazendo algo; não é uma "pesquisa" onde perguntamos coisas às pessoas) e velocidade (os dados estão imediatamente disponíveis na nuvem)
Brown (2014)	<i>Big data</i> é volume, velocidade, variedade. Então significa que você tem uma ou mais dessas três coisas.
Costa (2015)	<i>Big data</i> é um ecossistema, que se alimenta e retroalimenta de dados o tempo todo, de dados que estão digitalizados ou não e dos quais se têm a necessidade de tirar valor.
Taurion (2014)	Resumindo o que é big data em uma simples fórmula para conceitualizá-lo: Big Data = volume + variedade + velocidade + veracidade, tudo agregado + valor.
Moura (2015)	Teoricamente é aglutinar diversas informações em diversas fontes em um único depósito. Seria a manutenção, a orientação e o enriquecimento de um banco de dados de diversas fontes. Na prática isso é feito por meio de um depósito seja em nuvem, ou seja, de um depósito próprio que utiliza todas essas informações.
Guerreiro (2014)	<i>Big Data</i> se refere à necessidade de uma organização lidar com dados que, para serem analisados e gerarem resultados, demandam capacidade muito além daquela de que a empresa dispõe, seja com relação ao volume (mais comum), velocidade de tratamento ou variedade de formatos.
Lima Júnior (2014)	<i>Big data</i> é conjunto de dados (dataset) cujo tamanho está além da habilidade de ferramentas típicas de banco de dados em capturar, gerenciar e analisar.

Fonte: Autora

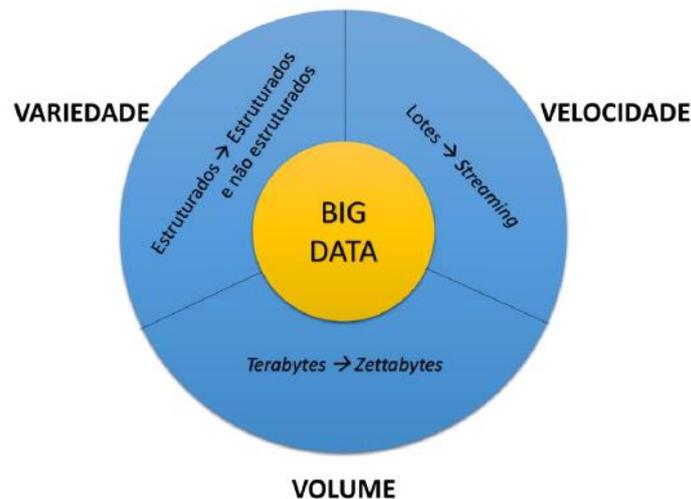
As discussões apresentadas em termos conceituais, remete esta pesquisa ao entendimento que um dos maiores problemas em relação ao *big data* está em se retirar dessas definições teóricas ou não, aplicações práticas, que evidenciem a importância do seu uso em qualquer que seja a área. *Big data* não é sobre uma única tecnologia, uma única arquitetura, ou um único caso de uso (OLOFSON, 2011). *Big data* na verdade não é somente tecnologia

ou um conjunto de softwares de análise de dados, é um ecossistema como aponta o americano Peter Hirshberg em uma entrevista para revista brasileira Exame³⁰ em 2013.

Estamos no início da era da “internet das coisas”. A internet está saindo do mundo virtual, das telas dos PCs, e tornando-se um elemento presente no mundo físico. Hoje os chips estão nos celulares, nos eletrodomésticos e nos carros, o que faz com que esses dispositivos possam ser conectados à internet. Essas conexões geram uma vasta quantidade de dados. A partir deles é possível analisar e entender, com mais precisão, o comportamento das pessoas (Trecho da entrevista de Peter Hirshberg para a revista Exame em 2013).

A partir disso, percebe-se entre as definições, que *big data* remete a três características principais, sendo referenciadas sempre como os 3Vs. Assim, de acordo com relatório da IBM (ZIKOPOULOS; EATON, 2012), as três características que definem o *big data* são volume, velocidade e variedade, pois juntas elas teriam criado a necessidade de novas habilidades e conhecimentos para melhorar a capacidade de lidar com as informações (Figura 7). Doug Laney foi o primeiro em 2011 a falar sobre os 3 Vs do *big data*, que serão vistos a seguir neste trabalho.

Figura 7: Três dimensões do *big data*



Fonte: Adaptada e traduzida de Zikopoulos e Eaton, 2012.

1. Volume: esta característica se refere à grande quantidade de dados digitais a serem analisados. Porém, não há uma definição precisa da medida, em bytes, da quantidade necessária para afirmar que alguma organização está lidando com um volume grande de dados. Para o analista e pesquisador da IDC (*International Data Corporation*), Carl W.

³⁰ Big data não faz milagres. Disponível em: <[dhttp://exame.abril.com.br/revista-exame/edicoes/1037/noticias/o-big-data-nao-faz-milagre](http://exame.abril.com.br/revista-exame/edicoes/1037/noticias/o-big-data-nao-faz-milagre)>. Acesso em: março 2014.

Olofson (OLOFSON *et al*, 2013) essa questão do volume depende do caso e da natureza do dado. “Algumas centenas de gigabytes podem muito bem caracterizar *big data* por causa da terceira dimensão, que é a da velocidade ou tempo necessário para processamento”. Assim, pode-se verificar que a questão da “grande quantidade de dados” é um aspecto relativo; depende de outros fatores.

Outros exemplos ilustram essa característica do *big data*. Os experimentos no *Large Hadron Colider* no CERN (Laboratório de Física de Partículas na Europa) geram 40 *terabytes* de dados por segundo, mais informação do que pode ser armazenada ou analisada pelas tecnologias atuais (muitos destes dados são simplesmente descartados, dada a incapacidade de armazená-los). Esta informação foi extraída da revista semanal *The Economist* de 2013, na qual é afirmado que exemplos de manuseio de grandes quantidades de informação como este também são encontrados em outros cenários, como nos bancos de dados do *Wallmart*, de tamanho estimado em torno de 2,5 *petabytes*, o equivalente a 167 vezes o conteúdo dos livros na biblioteca do Congresso Americano. Diante desse cenário, Xexéo (2013) observa que muitas vezes somos obrigados a escolher que tipos de dados queremos guardar. “O restante acabamos descartando. Cabe aí saber o que guardar. E guardar cada vez mais rápido é o desafio de hoje” (XEXÉO, 2013, p.19).

2.Velocidade: de acordo com o Gartner (BIG, 2013), velocidade significa tanto o quão rápido os dados estão sendo produzidos quanto o quão rápido os dados devem ser tratados para atender à demanda. Como exemplo têm-se as etiquetas RFID³¹ (*Radio-Frequency Identification*) e contadores inteligentes que estão impulsionando uma necessidade crescente de lidar com torrentes de dados em tempo quase real. Reagir rápido o suficiente para lidar com a velocidade é um desafio para a maioria das organizações. Assim, de acordo com Carl W. Olofson³² (2011) realizar a análise de uma grande massa de dados consumia uma hora de

³¹ Identificação por radiofrequência ou RFID é um método de identificação automática através de sinais de rádio, recuperando e armazenando dados remotamente através de dispositivos denominados etiquetas RFID. Uma etiqueta ou *tag* RFID é um transponder, pequeno objeto que pode ser colocado em uma pessoa, animal, equipamento, embalagem ou produto, dentre outros. Contém chips de silício e antenas que lhe permite responder aos sinais de rádio enviados por uma base transmissora. RFID utiliza transponders (os quais podem ser apenas lidos ou lidos e escritos) nos produtos, como uma alternativa aos códigos de barras, de modo a permitir a identificação do produto de alguma distância do scanner ou independente, fora de posicionamento. Tecnologia que viabiliza a comunicação de dados através de etiquetas com chips ou transponders que transmitem a informação a partir da passagem por um campo de indução (ex: muito usado em pedágio "sem parar"). (Wikipedia, 2014)

³² Artigo Big data: what it is and why you shouldcar. Disponível em:///C:/Users/Luciane/Downloads/IDC_Big%20Data_whitepaper_final.pdf. Acesso em: janeiro 2014.

processamento e, com o aperfeiçoamento desse processo, foi possível realizar em um segundo. Dessa forma, altera-se muito a perspectiva do que se pode “fazer com os resultados gerados, agregando mais valor ao dado”. Entende-se que avaliar isoladamente essas características não satisfaz a compreensão do que seja *big data*. Nesse aspecto, a velocidade no processamento é um fator fundamental para possibilitar a análise de um grande volume de dados no menor intervalo de tempo que seja possível.

3. Variedade: O grande volume de dados que temos atualmente é resultante da diversidade de informações. Assim, a variedade precisa ser lidada como um todo, talvez um determinado dado não seja útil se não for associado a outro. Os sistemas tradicionais são otimizados para processar dados que podem ser facilmente descritos na forma de tabelas, como uma planilha eletrônica, onde cada coluna tem tamanho constante ou previsíveis, mesmo que a quantidade de linhas seja muito grande. Muitos dos novos tipos de dados têm formatos mais livres (textos, imagens etc) ou com estruturas específicas (redes, por exemplo). (XEXÉO, 2013, p. 19).

À medida que o tema ganha discussões teóricas e práticas, outras duas características foram sendo adicionadas ao conceito que tratam do valor e veracidade (DEMCHENKO, 2013) e que para alguns autores são consideradas características também fundamentais do *big data* (TAURION, 2013; SCHÖNBERGER-MAYER; CUKIER, 2013).

4. Veracidade: Os dados precisam ser confiáveis, lidar com os outros fatores, como volume, velocidade e variedade pouco adianta se não forem reais. Assim sendo, é necessário que haja o máximo possível de consistência dos dados (TAURION, 2013).

5. Valor: Schönberger-Mayer e Cukier (2013) também discutem essa característica do *big data*, dizendo que o valor dos dados, no mundo do *big data*, passa do uso primário para o uso potencial no futuro, o que gera profundas consequências: afeta a forma como os negócios valorizam seus dados e a quem conferem acesso a eles. Segundo os autores, “isso permite, e talvez até obrigue, as empresas a mudar seus modelos de negócios e altera a forma como elas enxergam e utilizam os dados” (SCHÖNBERGER-MAYER; CUKIER, 2013, p. 70).

Esses autores chamam a atenção particularmente para as empresas de informação, como a Farecast ou o Google – nas quais, segundo eles, os fatos brutos entram num lado da linha de produção e saem do outro como informação processada. Com isso, os dados estão começando a parecer um novo elemento de produção. O valor imediato da maioria dos dados

é evidente para os que o coletam. Por exemplo: as lojas registram dados de vendas para realizar um adequado balanço financeiro. As fábricas monitoram a produção para garantir que esteja adequada à demanda e a seus padrões de qualidade. Os sites registram cada clique dos usuários para analisar e otimizar o conteúdo que fornecem aos visitantes.

Ao contrário das coisas materiais, o valor dos dados não diminui com o uso; ele pode ser continuamente reprocessado, reagregando valor a cada novo uso. “Os dados são o que os economistas chamam de bem “sem concorrente”: seu uso por uma pessoa não impede o uso por outra, e a informação não se desgasta como as coisas materiais”. (SCHÖNBERGER-MAYER; CUKIER, 2013, p 71). Como os dados podem ser usados várias vezes com o mesmo objetivo, o mais importante é que podem ser utilizados com vários objetivos também. Da mesma forma, César Taurion (2014) observa que, a maioria das empresas usa o dado para efetuar e gerenciar uma transação, mas não valoriza seu potencial de reutilização.

Uma imagem que podemos associar a esta visão é a de um iceberg, onde apenas vemos sua parte visível, muito pequena, que é o dado usado na transação, mas não vemos sua imensa parte submersa, que é a sua reutilização de diversas formas. O dado deve ser valorizado não apenas pelo seu valor imediato, transacional, mas pelo seu valor futuro, quando utilizado em análises preditivas, identificação de correlações etc. (TAURION, 2014)

Um exemplo citado pelo autor são os dados gerados pelas empresas telefônicas. Segundo Taurion (2014), cada ligação gera um ou mais registros que são usados basicamente para faturamento e análise de desempenho da rede, como reposicionar as estações radio-base dos celulares. Depois de cumprirem suas funções são descartados. “Mas, estes dados podem ter muito valor para outros usos, como análises preditivas de movimentações de clientes para varejistas, impactos de campanha de marketing e assim por diante” (TAURION, 2014). O autor observa que um caso que vale a pena ser estudado é a empresa criada pela operadora Telefonica, na Espanha, que criou uma empresa separada, Telefonica *Dynamic Insights* para vender dados de localização e seus assinantes (devidamente anonimizados) obtidos naturalmente pelos seus sistemas transacionais de gestão de redes. Taurion cita também exemplo de como o dado pode ser usado a partir das câmeras de vigilância. Os dados, segundo ele, podem ser usados de forma não prevista inicialmente.

Imaginemos, por exemplo, que um varejista colocou dezenas de câmeras de vigilância, em sua loja, preocupado com furtos de mercadorias. Mas, se além de se precaver contra furtos estas imagens puderem ser analisadas e usadas para avaliar a movimentação dos clientes pelos corredores da loja e em quais gôndolas elas param mais tempo? Poderiam checar se os locais de promoções realmente atraem os clientes. Não teria um valor significativo para o varejista? Assim, ao invés de apenas reduzir custos, diminuindo os furtos, o sistema levaria a um aumento de receitas, melhorando a disposição dos produtos pelos corredores. (TAURION, 2014)

Com esse exemplo é possível entender que a coleta de dados que tinha uma função primária passou para uma função secundária. Os dados foram reutilizados para outra finalidade prevista ou não. E nessa perspectiva que o *big data* se coloca como uma característica de valor.

3.2.2 Utilidade do *big data*

Mas afinal para que serve o *big data*? Esse é um dos principais questionamentos que tanto o mercado, quanto a presente investigação faz. Pensemos aqui as possibilidades do uso do *big data*, principalmente do ponto de vista estratégico e de negócios a partir da quinta característica do *big data*, que é o valor. Segundo os especialistas da IBM³³ é possível com o *big data*:

- **Identificação de padrões:** que podem auxiliar na tomada de decisões. Por exemplo, a percepção de que um grande número das pessoas sofre de um problema comum gera o insight em relação às carências que as empresas podem preencher.
- **Personalização:** o armazenamento de dados relativos aos hábitos e interações das pessoas permite que elas tenham atenção personalizada, o que torna o atendimento e a resolução dos seus problemas muito mais efetiva.
- **Planejamento estratégico:** as empresas podem planejar suas ações com base em informações previamente mapeadas direcionando suas decisões com base em dados.
- **Inovação:** o uso de *big data* possibilita um conhecimento maior dos clientes e do mercado como um todo, facilitando o processo de criação de novos produtos e serviços.

Assim, o *big data* torna-se viável e rentável, emergindo formas de “domar” dados massivos, pois, dentro desses dados existem padrões e informações valiosas, previamente ocultos devido à sua extração complexa. Da mesma forma para a Fundação TechAmerica (2012) o *big data* também pode ser utilizado como estratégias de governos, trazendo benefícios para saúde, educação, transporte e também ajuda na detecção de crimes em relação ao imposto de renda, e também a crimes digitais.

Um exemplo é a forma como o governo brasileiro vem usando o *big data*. Ele tem ajudado o poder público a gastar de forma mais eficiente. Foi desenvolvido um sistema inteligente de compras públicas para ajudar na otimização de gastos. O projeto *big data* foi elaborado pela Fundação Getúlio Vargas (FGV). O sistema foi debatido e apresentado pela

³³Informações retiradas do site da IBM.

primeira vez em 2013 na França durante o Fórum da Organização para Cooperação e Desenvolvimento (CORRÊA, 2013).

Apresenta-se um Quadro (3) de mudanças elaborado pela empresa ELife apontando de forma resumida as possibilidades trazidas com a aplicação do *big data*.

Quadro 3: Mudanças do *big data*

ANTES	AGORA
Amostra de dados	N=all. Hoje é possível analisar (quase tudo) que é gerado pelas mídias sociais, sem recorrer a amostras.
Silos de dados	Fim dos silos de dados. Dados de diferentes mídias sociais podem ser armazenados em um único armazém de dados.
Dados já utilizados uma única vez podem ser descartados	A informação pode ser reusada. O custo de armazenamento de dados na nuvem permite que você salve um grande volume de dados e re-use-o, mesmo que os dados tenham sido coletados com outro propósito.
Porque?	“O quê” é melhor que “Porquê?” para social big data. Geralmente não vamos conseguir ter explicações para tudo, mas teremos mais certeza sobre o quê estamos falando graças ao grande volume de dados.
Versão única da verdade	O fim de uma versão única da verdade. Os dados contam histórias e as histórias dependem das questões perguntadas. Em outras palavras acostume-se a ter muitas versões da verdade, dependendo das perguntas que fizer.
Dados estruturados em bancos de dados tradicionais	A natureza dos dados de social é desestruturada, o que significa que há poucos ou nenhum campo estruturado. Apenas uma corrente contínua de texto. Os analistas devem aprender a lidar com a desestruturação e extrair os dados destes bancos desestruturados, limpando o lixo, através de software específico para este fim.
Relatórios consolidados	Os relatórios consolidados, geralmente uma fotografia da realidade, estão sendo pouco a pouco substituídos por dashboards em tempo real. A própria E.life oferece dashboards em real time do social big data.

Fonte: ELife (2013)

O quadro comparativo elaborado pela empresa ELife (2014) dá uma clara visão das transformações antes e depois do *big data*. Assim, segundo Schönberger-Mayer e Cukier (2013) o *big data* se relaciona com três importantes mudanças de mentalidade interligadas que se reforçam. A primeira coisa a ser observada com o *big data* é a capacidade de analisar grandes quantidades de dados sobre um tema em vez de ser obrigada a contar com conjuntos menores. A segunda mudança é a disposição de aceitar a real confusão dos dados em vez de privilegiar a exatidão e a terceira seria o maior respeito por correlações do que pela contínua busca pela causalidade elusiva (SCHÖNBERG-MAYER; CUKIER, 2013, p.13).

3.2.3 De onde vêm os dados?

Xexéo (2013) aponta que os dados estão presentes no comércio e na indústria. Segundo o autor, as informações são geradas a partir de cada venda em uma rede de supermercados, por exemplo. E se essas informações forem cruzadas com mensagens em

redes sociais sobre mercados, produtos, receitas e notícias na mídia ou com as regiões geográficas, em que acontecem, os sistemas que analisam toda essa informação permitem aos donos dessas lojas obterem tanto dados de perfis de consumo individualizado quanto da experiência de compra do cliente e seus gastos.

Já na indústria, Xexéo (2013) destaca que um dos fatores que aumenta a quantidade de dados é a multiplicação dos sensores de vários tipos, de câmeras de alta definição a simples contadores ou termômetros. O autor exemplifica, dizendo que uma empresa pode colocar sensores não só em sua fábrica para analisar o processo produtivo, mas também em seus produtos, que enviam ao fabricante, dados sobre sua localização geográfica, seu uso e desempenho.

Zeferino (2013) aponta a evolução histórica do *big data*, a partir de três níveis fundamentais de produção de dados:

- A produção de dados em regime de ***Input*** - Situação que retrata os primórdios dos primeiros computadores pessoais, sobre os quais existia a necessidade de armazenar dados diretamente em cada PC, e que marcou uma viragem determinante na vida tecnológica das empresas.
- A produção de dados em regime de ***User Generated Content*** (UGC) – Situação que abrange a realidade da chamada Web 2.0, na qual se evidencia o papel preponderante do próprio utilizador na criação e geração de conteúdos em regime colaborativo e de partilha.
- A produção de dados em regime de ***Machine-Generated Data*** (MGD) – Situação que retrata o processamento automatizado de dados sem intervenção humana e que se refere ao funcionamento de complexos sistemas integrados em ambiente de computação contínua.

No quadro 4 são apresentadas as principais fontes geradoras de dados.

Quadro 4: Origem dos dados

Pessoais	Financeiros	Trabalho	Redes Sociais	Comportamento	Pervasivos
RG/CPF	Operações bancárias	Bilhete único	Facebook	Compras	Radar de trânsito
Imposto de Renda	Cartão de crédito	Ponto	Twitter	Recomendações	GPS
Previdência	Serasa/SPC	Login		Pesquisa	Bluetooth
Plano de Saúde	Financiamento/Consórcio	Vale alimentação		Cookies	WiFi
				Canais	RFID

Os dados para compor o *big data* podem vir de diversas fontes e formatos, desde dados de documentos pessoais até dados de radares de trânsito. A maioria das coisas hoje, até aqueles equipamentos residenciais, já oferecem informações e dados sobre seus estados. Geladeiras, por exemplo, ligadas à Internet oferecem dados sobre o estado e ritmo de consumo dos produtos que estão guardadas nelas. Esse fenômeno é identificado por muitos pesquisadores como a “Internet das coisas” (IoT) (LEMOS, 2013, p.239). A internet das coisas é um conjunto de redes, sensores, atuadores, objetos ligados por sistemas informatizados que ampliam a comunicação entre pessoas e objetos (o sensor no carro avisando a hora da revisão, por exemplo) e entre os objetos de forma autônoma, automática e sensível ao contexto (o sensor do carro alertando sobre acidentes no caminho).

Ainda sobre os dados, Xexéo (2013) observa que no Brasil, em especial, a indústria de petróleo é grande criadora de dados, que vão da pesquisa sísmica inicial e monitoração eletrônica de poços até a venda de combustível na bomba dos postos de abastecimento. Os dados dos bancos de dados da Agência Nacional do Petróleo (ANP) são usados por várias empresas de petróleo, que produzem ainda mais informações. A partir deles, são gerados imagens e filmes que são analisados e geram textos e relatórios, que, por sua vez, geram investimentos financeiros, movimentos da bolsa e notícias na mídia. Assim, entende-se que os dados gerados circulam por diversos ambientes e diversos formatos o tempo todo.

A geração dessa massa informacional está em todo lugar que vai desde as interações da web, na própria busca no Google até a compra da passagem aérea (ARRIGONI, 2013). No entanto, nem todos os dados são necessariamente úteis para análise de *big data*. Segundo as perspectivas dos analistas de dados David Reinsel e John Gantz (2012), do IDC Group

apresentadas no site do grupo³⁴ somente alguns dados são tidos como maduros para as análises de *big data*.

Para esses analistas, destacam-se as **imagens de vigilância**, que fornecem metadados como data, hora, local, automaticamente ligados a um arquivo de vídeo. Hoje as câmeras podem fornecer imagens que são capturadas e analisadas em tempo real. Reinsel e Gantz (2012) observam que esse tipo de informação pode ajudar a agilizar investigações de crimes e melhorar a análise de varejo para os padrões de tráfego de consumo. **Os dispositivos médicos** incorporados no corpo humano também contribuem para as análises de dados na área de saúde, podendo trazer muitos benefícios para a pesquisa e para o combate de doenças e vírus. **Imagens de consumo** (fotografias de indivíduos) que são compartilhados em ambientes de redes sociais digitais, segundo os analistas podem ajudar a identificar tipos de consumidores e falar muito sobre as pessoas e suas características e por fim os dados da **mídia social**, que hoje é considerada uma grande fonte de grandes volumes de dados para trazer insights para o mercado ou mesmo para ajudar a prever uma eleição.

Isaca (2013) também aponta os tipos de dados pessoais que são coletados para o *big data*. São eles:

Dado voluntário: criado e explicitamente compartilhado por indivíduos. Por exemplo, perfis de redes sociais;

Dado observado: capturado ao gravar ações do indivíduo. Por exemplo, dados de localização no uso de telefones celulares;

Dado inferido: dado sobre indivíduos com base em informações voluntárias ou observadas. Por exemplo, classificação de crédito.

Outro autor também referencia os dados de acordo com os seus tipos. Ohl (2014) classifica os dados como implícitos e explícitos.

Os dados **implícitos** são aqueles coletados sem que necessariamente se tenha a anuência das pessoas durante um processo de análise. Esses dados podem ser desde a compra de um produto, até um compartilhamento ou comentário de postagem em blogs ou redes sociais entre amigos. Ohl (2014)³⁵ observa que por esse fato pode carregar consigo uma conotação sombria e muitas vezes apelidada de “*Big Brother*”, “*Big Oi*” e “*Big Government*”. Segundo

³⁴ Disponível em: <http://www.emc.com/leadership/digital-universe/iview/index.htm>. Acesso em: dezembro de 2014.

³⁵ Informações retiradas do artigo “Big Data: como analisar informações com qualidade”. Disponível em: <http://corporate.canaltech.com.br/coluna/big-data/Big-Data-como-analisar-informacoes-com-qualidade/>. Acesso em: Dezembro 2014.

Ohl (2014) com estes dados empresas, governos, organizações podem conhecer os hábitos de seus clientes/cidadãos e, desta forma, muitas vezes conseguem prever suas próximas ações.

Cada bloco de dado disponível está sendo destrinchado e esmiuçado para posterior análise. Os maiores varejistas, das cadeias de supermercados aos bancos de investimentos, têm uma área de “análise preditiva”, focada não apenas em entender os hábitos de compras dos consumidores, mas também seus hábitos pessoais, buscando assim uma forma mais eficiente de comunicar e vender para eles (OHL, 2014).

No entanto, Ohl (2014) ressalta que, apesar de a coleta de dados desse tipo ser a mais utilizada em se tratando de *big data*, o analista corre o risco de ter predições erradas, pelo fato de que as informações podem ser obtidas em grande escala. Isso porque eles são baseados na coleta passiva dos hábitos e comportamentos das pessoas. “E é exatamente por isso que não são 100% à prova de falhas” (OHL, 2014). E ele explica:

Com esse tipo de dados, não é possível para qualquer varejista saber se uma avó está comprando um presente de aniversário para seu neto ou para si própria. Da mesma forma, este varejista não consegue saber se você está comprando um livro para você ou como presente para um amigo. E, independentemente do quão genial seja o analista, ele nunca conseguirá sugerir uma música certa sem PERGUNTAR para a pessoa se determinado ritmo lhe agrada (OHL, 2014).

Então, para que as respostas sejam mais precisas o importante é fazer perguntas para o indivíduo. Quando se usa a pergunta para a coleta de dados isso é chamado de dado **explícito**. Segundo Ohl (2014) historicamente, dados explícitos custam caro e demandam muito tempo para serem apurados. Para ele estes são os motivos por que tradicionalmente esses dados acabam recebendo uma grande relevância nas análises de *big data*. Entretanto, Ohl (2014) chama a atenção que a tecnologia tem mudado isto. A internet permite que as empresas obtenham dados explícitos em grande escala, por meio de uma variedade de plataformas. Esses tipos de dados são obtidos através dos métodos de coleta mais tradicionais, como pesquisas de opinião/ satisfação; ou por meio de resenhas publicadas na internet, formulários de sugestões preenchidos *on* e *off-line*, curtidas em uma postagem ou em um perfil de rede social ou blog. Esses dados se configuram assim, porque a maioria das pessoas usuárias de ambientes da Internet tem conhecimento, mesmo de maneira superficial, de que suas ações podem estar sendo monitoradas e capturadas por empresas para diversas finalidades.

Assim, o importante em estratégias com o *big data* é sempre utilizar os dois tipos de dados (implícitos e explícitos), porque eles se complementam. Enquanto os dados implícitos querem saber “o quê?”, os dados explícitos querem saber o “por quê?” das ações e reações de uma pessoa, ou massa de pessoas. Fazendo a coleta desses dois tipos de dados, e interpretando-os em conjunto e com cuidado, é possível fazer boas previsões e ter um

conhecimento dos sentimentos do seu consumidor quanto à sua marca, ao seu produto ou a uma campanha específica. Assim, também é possível descobrir novas tendências e ter *insights* para novos nichos e negócios.

No entanto, independente do tipo de dado que se quer, para alguns especialistas o *big data* ainda passa por muitos desafios que devem ser levados em consideração. Na visão de Simon Szykman³⁶, diretor de informática do Departamento de Comércio Norte-americano, esses desafios podem ser divididos em: como adquirir; armazenar; processar; transmitir e disseminar; gerenciar e manter; arquivar por longo prazo; garantir a segurança; treinar pessoas para usá-los; pagar por tudo isso. Além disso, a preocupação está em como entender os dados de forma isolada e como relacioná-los.

Para isso, a sociedade precisa de novas teorias e tecnologias. Uma das principais aplicações de ter à disposição uma grande quantidade de dados é descobrir algo novo. Isso, segundo Xexéo (2013), pode se dar por detecções de padrões, levantamento de regras ou aprendizagem de máquina. As detecções de padrões têm por objetivo encontrar em grandes volumes de dados padrões de comportamentos e isso é feito a partir da mineração de dados.

Já a aprendizagem de máquina, termo que vem do inglês *Machine Learning*, é a área de Inteligência Artificial que tem por objetivo o desenvolvimento de técnicas computacionais sobre processo de aprendizado (BISHOP, 2007). Em síntese, é uma área de sistemas inteligentes que trata do desenvolvimento de algoritmos e técnicas que permitem computadores “aprender”.

3.3 USOS PRÁTICOS DO *BIG DATA*

Hoje, várias áreas são guiadas por programas desenvolvidos para análise de informações a partir do emprego do *big data*. Para Schönberger- Mayer e Cukier (2013), os usos do *big data* se aplicam tanto ao setor público quanto às empresas privadas, com destaque também para as áreas de negócios, tecnologia, saúde, cidades inteligentes (GARTNER, 2013). Nesta parte do trabalho, apresentam-se algumas áreas consideradas relevantes para o entendimento das possibilidades geradas com o *big data* a partir dos seus usos e aplicações práticas. Aqui são destacados os usos feitos por governos nas suas administrações e também com a finalidade de espionagem e controle, considerado um ponto crítico no uso de análise de grandes volumes de dados com o propósito de vigiar, controlar e mesmo punir as pessoas.

³⁶ Informações retiradas do artigo “Big Data computação para uma sociedade conectada e digitalizada”, Geraldo Xexéo. Revista Ciência Hoje, v. 51, n 306, p. 18-23. Agosto 2013.

3.3.1 *Big data* e o Governo

Em Nova York, na última década, de 2002 a 2013, o prefeito da cidade, Michael Bloomberg, implementou um grande armazém de dados para a cidade. O objetivo era aperfeiçoar decisões da gestão pública, como quais imóveis – entre os 900 mil da cidade – vistoriar para impedir a ocorrência de grandes incêndios. O primeiro passo foi identificar as diferentes fontes de dados que geram informações sobre a cidade: de imóveis alugados, passando por emergências médicas, reclamações sobre barulho na vizinhança, denúncias de crimes, até avistamento de ratos.

Os dados coletados vinham de diversas fontes. Estes dados estavam desestruturados e em silos diferentes (diferentes bases de dados de acordo com a esfera municipal que tratava do problema). Primeiro passaram por um processo de estruturação para que fosse possível sua manipulação. Foi criado então um modelo estatístico para manipular os dados. A partir da implementação do modelo, cada nova queixa recebida pela prefeitura para perigo de incêndio passou a ser checada usando o algoritmo do novo modelo. Resultado desse trabalho apresentou que o índice de queixas atendidas que de fato representavam algum perigo eminente saltaram de 30% para 70%. A assertividade nas previsões economizou recursos para a cidade de Nova Iorque, uma vez que evitava o deslocamento de equipes para alarmes falsos. Isso também foi aplicado a bueiros que explodiam em Nova Iorque. Esse mesmo problema, que tem assolado outras cidades no mundo como o Rio de Janeiro, na Big Apple também foi tratado com *big data*. Para descobrir bueiros a ponto de explodir, foi cruzada uma grande massa de dados de diferentes fontes, como mapas de tubulações de gás, telefone, eletricidade, além do trânsito nas imediações. Esses dois exemplos ilustram o uso e análise dos grandes dados no mercado.

Outro exemplo de aplicação do *big data* citada por Schönberger-Mayer e Cukier (2013) são as “cidades inteligentes” (*Smart Cities*), que, atualmente, segundo os autores, estão sendo planejadas em diversos países e visualizam para um futuro próximo espaços urbanos altamente intensivos no uso de tecnologias de informação e comunicação. Sensores disseminados em equipamentos de infraestrutura urbana, informações produzidas por dispositivos pessoais de comunicação, redes sem fio, aplicativos (app) inteligentes, *posts* em mídias sociais baseadas em geolocalização, serviços de computação em nuvem. Todos esses elementos vislumbrados como parte essencial das cidades inteligentes – pelo menos do ponto de vista tecnocrático das grandes empresas de tecnologia da informação – poderiam produzir e alimentar grandes bases de dados, com o objetivo de reagir de forma rápida a demandas

individuais e mudanças no ambiente urbano. A partir dessa perspectiva em uma cidade inteligente, as tecnologias de comunicação e de informação, bem como os sistemas de sensores, desaparecem à medida que se tornam embutidas nos objetivos físicos e nos ambientes (STEVENTON; WRIGHT, 2006).

O termo “cidades inteligentes” é oriundo de “cidades digitais”, muito discutidas nos anos de 1990. Digital era compreendido como acesso a computadores e a implantação da Internet no espaço urbano. Já inteligente refere-se a processos informatizados sensíveis ao contexto, lidando com um gigantesco volume de dados, redes em nuvens e comunicação autônoma. O projeto de cidades inteligentes³⁷, além de integrar diferentes âmbitos e tecnologias, tem como principal atrativo e diferencial a análise de dados obtidos em grande escala do comportamento urbano do cidadão, mediante a integração dos grandes dados digitais, prometendo melhorar e definir os serviços oferecidos, em áreas como transporte, energia, meio ambiente e comunicação³⁸.

Municípios como Songalo, na Coreia do Sul, e Masdar, em Dubai, estão sendo construídos, literalmente, desde a fundação, com quase toda a infraestrutura conectada à Internet³⁹.

É importante destacar os projetos que estão em andamento também no Brasil. Em setembro de 2013, foi anunciado que o *big data* iria monitorar a população por meio do registro celular. A Fundação Getúlio Vargas (FGV) desenvolveu um projeto de *big data* para o monitoramento da população da Região Metropolitana do Rio de Janeiro a partir dos CDR das ligações de usuários das operadoras celulares. Segundo Alexandre G. Evsukoff, professor da FGV, o projeto tem por objetivo identificar os padrões de mobilidade de um universo de 2 milhões de pessoas e é utilizado para o desenvolvimento de políticas públicas e planejamento na área de transporte e controle de epidemias⁴⁰. O projeto conta com apoio da Fundação de Amparo à Pesquisa do Rio de Janeiro (Faperj) e parceria com a Coppe-UFRJ e o Massachusetts Institute of Technology (MIT).

³⁷As *smart cities* do futuro partem das pessoas e das comunidades onde vivem e trabalham. É um novo paradigma na forma de fazer cidades, que exige repensar estratégias, tecnologias, modelos e processos urbanos para responder aos atuais desafios ligados à qualidade de vida, ao equilíbrio do ambiente e eficiência dos recursos naturais, às desigualdades e à exclusão social.

³⁸ Revista IMasters. Disponível em < <http://issuu.com/imasters/docs/revistaimasters5>>. Acesso em: 02 de Dezembro de 2014.

³⁹ Informações retiradas da reportagem Asia constrói cidades inteligentes “do zero” ao custo de US\$ 102bi. Disponível em: <http://www1.folha.uol.com.br/fsp/tec/116712-asia-constro-cidades-inteligentes-do-zero-ao-custo-de-us-102-bi.shtml>. Acesso em: 26 set.2014.

⁴⁰ O Globo. Disponível em:<http://oglobo.globo.com/projetos/tirio_2013/mat2.asp>. Acesso em: 02 de Outubro de 2014.

No Rio de Janeiro, a prefeitura faz investimentos na área de tecnologia da informação na captura e análise de dados dos cidadãos por meio do Centro de Operações Rio – COR -, que foi construído no governo de Eduardo Paes (PMDB) por exigência do Comitê Olímpico Internacional, para a cidade sediar os jogos de 2016. A IBM foi contratada para elaborar a tecnologia do COR, cuja intenção é centralizar informações dispersas em diferentes órgãos públicos e sensores de monitoramento espalhados pela cidade como câmeras de trânsito e pluviômetros, para possibilitar, segundo a IBM, a visualização integrada de dados e agilizar a resposta da prefeitura a problemas comuns da cidade, como enchentes, deslizamentos, bloqueios no tráfego e emergências policiais.

Além disso, desde 2013, é feito monitoramento em redes sociais digitais para identificar o sentimento dos cariocas em relação ao governo municipal. As ferramentas de monitoramento foram desenvolvidas pela SAP, outra multinacional de tecnologia de informação. Esse tipo de pesquisa é, habitualmente, empreendida por profissionais de marketing para avaliar o desempenho de produtos ou a imagem de políticos e programas de governo. Além disso, existe um grupo chamado Pensa – Sala de Ideias⁴¹, que, desde 2013, a partir das análises de dados e informações dos cidadãos cariocas e da administração municipal, apresenta projetos e propostas para aperfeiçoar os serviços para a cidade. O coordenador do grupo e cientista de dados Pablo Cerdeiro⁴² diz que a preocupação desse projeto é centralizar dados dos diversos setores e órgãos e disso tirar ideias e projetos para ajudar na gestão do governo municipal. Cerdeiro conta que são utilizadas as informações geradas pelos mais de 6,3 milhões de cidadãos cariocas e os mais variados sensores. Os dados vêm de índices obtidos pelas diversas secretarias e órgãos públicos, somados aos que chegam via câmeras de controle, GPS, radares e canais como a central telefônica e mobile de atendimento ao público 1746, além de redes sociais. Essas fontes chegam a gerar mais de um milhão de dados da cidade por dia. A partir das análises e cruzamentos de dados da cidade, são gerados relatórios diários sobre as reais condições da cidade em diversos setores.

Desde a sua criação até hoje o grupo já desenvolveu vários projetos considerados importantes por Cerdeiro⁴³. O primeiro objeto de estudo do grupo Pensa foi o mapeamento das árvores da cidade. A partir da informação de que elas ficam vulneráveis com ventos superiores a 40 km/h, o Centro de Operações da Prefeitura passou a enviar sistemas de alerta para a Comlurb

⁴¹ O projeto é formado por um grupo multidisciplinar de servidores públicos que ganharam destaque em projetos de liderança e doutores e pós-graduados em análise de dados e ciência da computação. O projeto é coordenado pelo advogado, professor da Fundação Getúlio Vargas e cientista de dados, Pablo Cerdeiro.

⁴² Palestra proferida no Primeiro Encontro de Big Data, na Coppe UFRJ em 2014.

⁴³ Disponível em: <http://vejario.abril.com.br/materia/cidade/rio-tem-tropa-elite-big-data>. Acesso em: 20 mar. 2015.

– que é responsável pela limpeza da cidade – toda vez que os ventos chegavam nesse índice. O órgão, a partir desses estudos, 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 frequentemente afetados. Mas, além desse, foram feitos projetos na área da saúde para prevenção e combate à dengue. As ações realizadas a partir das análises de *big data* reduziram em 94% o número de casos de dengue na cidade no ano de 2014 ao identificar correlações que apontavam as regiões mais afetadas pela Dengue.

No que se refere à mobilidade urbana, a prefeitura tem uma parceria com o aplicativo *Waze*⁴⁴. O serviço informa todos os tipos de ocorrências no trânsito desde engarrafamento até acidentes, indicando também os melhores caminhos e trajetos para os motoristas chegarem aos seus destinos. No Rio, a ferramenta funciona como uma camada de dados adicional às informações reunidas no Centro de Operações⁴⁵. Os dados ajudam a administração a direcionar suas ações, como consertar um semáforo cujo problema foi apontado por usuários do aplicativo. Além disso, a prefeitura do Rio⁴⁶ exibe em tempo real em painéis eletrônicos, distribuídos pelos principais pontos das vias, informações do tempo que o motorista leva para percorrer determinada localidade na cidade. A imagem 1 exemplifica esse caso. O painel digital *big data* do Waze faz parte do programa *Connected Citizens* (Cidades Conectadas)

⁴⁴ Waze é um aplicativo de trânsito e navegação para celular baseado em uma comunidade. Os motoristas compartilham informações de trânsito das vias em tempo real. O aplicativo nasceu no Brasil, mas já se espalhou por outros países. Rio de Janeiro, San Jose, Barcelona, Jacarta, Tel Aviv, Boston, estado de Utah, estado da Flórida, Los Angeles e a Polícia de Nova York já utilizam o Programa. Waze concede informações como velocidade média dos veículos em uma via e o grau de lentidão minuto a minuto, tempo médio de um percurso e atrasos provocados por congestionamentos; rotas alternativas; acidentes automobilísticos e incidentes na via como buracos. Em contrapartida, a prefeitura do Rio informa as vias bloqueadas e obras que atrapalhem o trânsito. Em outubro de 2014 eram mais de 50 mil usuários do aplicativo na cidade do Rio de Janeiro.

⁴⁶ A cidade é a primeira no Brasil a integrar o Waze aos painéis de sinalização.

Imagem 1: Painéis de Comunicação Digital Big Data



Fonte: Divulgação Prefeitura/G1

Ao contrário de exibir um tempo de deslocamento estimado, o sistema informa o tempo do percurso baseado nos dados dos usuários, sem fazer um julgamento subjetivo como bom, intenso ou lento. O cálculo dos tempos de percurso é processado por ferramenta do Google, que leva em conta dados do trânsito monitorado em tempo real. Os painéis estão distribuídos em pontos da cidade com maior fluxo de veículos.

3.3.2 Big data e espionagem

Em 29 de março de 2012, o governo dos EUA lançou a *big data Research and Development Initiative*, dotada de um orçamento inicial de 200 milhões de dólares, com o objetivo de acelerar o ritmo do desenvolvimento da ciência e tecnologia, de fortalecer a segurança nacional e de transformar o processo de ensino e aprendizagem. Mas antes dessa iniciativa o governo dos EUA já utilizava o *big data* para atividades de espionagem à escala global, mas de forma clandestina⁴⁷.

O projeto PRISM, conduzido pela National Security Agency (NSA), funcionou desde 2007 de forma secreta até junho de 2013, momento em que os jornais britânicos Guardian e norte-americano Washington Post revelaram as atividades daquela agência, com base em documentos e depoimentos de Edward Snowden, um ex-analista contratado pela NSA que abandonou os EUA e se exilou na Rússia. O PRISM é capaz de registrar os conteúdos de todas as formas de comunicação eletrônica em âmbito mundial, incluindo todos os e-mails, todas as chamadas de celulares, todas as navegações feitas na Internet, todas as imagens e dados obtidos em satélites, bem como toda espécie de dados pessoais, empresariais, institucionais e governamentais que estejam sob forma digital e circulem nas redes de telecomunicações. Segundo Washington Post, o que é preocupante sobre o PRISM não é a

⁴⁷ Informações retiradas do site. Disponível em: http://issuu.com/telaviva/docs/ti_93/9?e=1344148/4817402. Acesso em: 08 de fevereiro de 2015.

coleta de dados, mas sim o tipo de dado coletado⁴⁸. Porém, como aponta Bruno (2013), o governo americano, desde os atentados de 11 de setembro de 2001, numa série de programas e leis, vem criando zonas de exceção que autorizam o monitoramento de dados e comunicações da população civil, sob a justificativa do combate ao terrorismo (Patriot Act, 2001 e 2006; Waranteless Domestic Surveillance Program, 2001; Total Information Awareness Program, 2002; Foreign Intelligence Surveillance Court, 2007). No entanto, uma das peculiaridades do PRISM, ativo desde 2007, seria a sua aliança com as corporações privadas na internet, implicando uma mudança não só quantitativa, mas também qualitativa no acesso aos dados pessoais de internautas, uma vez que, segundo os documentos, a NSA teria acesso não simplesmente aos metadados das navegações (horário e local de conexão, por exemplo), mas também aos seus conteúdos (histórico de buscas, conteúdo de e-mails, transferência de arquivos etc.)⁴⁹ (BRUNO, 2013)

Entre os países “espionados” pela agência americana estava o Brasil. A base de espionagem da NSA e da CIA era Brasília, que faziam interceptações de comunicações do governo e dos brasileiros. Os equipamentos tinham como objetivos captar dados em linhas de comunicação no Brasil e usariam diversas ferramentas de software para obter as informações. Entre elas estava o PRISM, para acesso a e conversas on-line; Boundless Informant, que determinaria quem enviou o que para quem e quando; X-Keyscore, que faria reconhecimento de textos em mensagens; Fairview que é capaz de monitorar grandes quantidades de dados. Tudo era captado em tempo real. Segundo James Bamford, que é autor de vários livros sobre a NSA, a rede da agência americana inclui salas de espionagem em operadoras de telecomunicações e um grande data center em Bluffdale, no estado americano de Utah. Bamford, em entrevista para o jornal O Globo, diz que com esse aparato, basta um agente digitar o nome de uma pessoa num computador para que todas as conversas dela passem a ser registradas e todos os e-mails e mensagens instantâneas passem a ser interceptados. Depois desse processo, os computadores se encarregam de analisar esses dados e relatórios são gerados dando dimensões das capturas e análises dos dados.

O governo americano também mantinha bases de vigilância e espionagem em outros países. O *The Guardian*⁵⁰ revelou que as autoridades dos serviços secretos britânicos

⁴⁸ Disponível em: < <http://gizmodo.uol.com.br/o-que-e-prism/>>. Acesso em: 20 de março de 2015.

⁴⁹ Informações retiradas do O Globo on line. Disponível em: <http://oglobo.globo.com/blogs/prosa/posts/2013/06/22/internet-territorio-privilegiado-da-vigilancia-de-massa-500790.asp>. Acesso em: 02 de março de 2015.

⁵⁰ Informações retiradas da reportagem do site The Guardian. Disponível em: <http://www.theguardian.com/world/2013/nov/20/us-uk-secret-deal-surveillance-personal-data>. Acesso em: 06 de março de 2015.

aprovaram o pedido americano para analisarem e armazenarem registros de chamadas telefônicas, Internet e de cidadãos do Reino Unido. Essa vigilância incluía pessoas que não eram suspeitas, de qualquer atividade criminal, segundo revelava um documento divulgado pelo ex-funcionário da National Security Agency (NSA) Edward Snowden. As informações coletadas estão sendo colocadas em bancos de dados, onde pode ser disponibilizadas a outros membros da inteligência dos Estados Unidos e da comunidade militar.

Mas a prática da espionagem a partir da análise de dados e informações das pessoas não é algo novo. Até bem pouco antes da queda do Muro de Berlim (1989), a Polícia Secreta da Alemanha Oriental (Stasi) espionava milhões de pessoas. Observava ruas, carros, casas, abriam cartas, espionavam contas bancárias, ligações telefônicas e infiltrava agentes em empresas e famílias. Com pelo menos 40 milhões de fichas e mais de 100 quilômetros de documentos, gravavam e detalhavam os aspectos mais íntimos das pessoas.

Dessa forma, *big data* ganha diferentes discussões a partir dos diversos campos em que pode ser aplicado, expondo as suas possibilidades e limites de uso. O importante é entender que mesmo sob uma perspectiva otimista dos seus usos no campo dos negócios ou mesmo nas gerências de cidades e governos, *big data* sozinho não faz milagres em nenhum projeto. O seu uso representa um desafio tecnológico, pois demanda atenção à infraestrutura e tecnologias analíticas (TAURION, 2014, p. 123), bem como têm necessidades de profissionais capacitados tanto para gerenciar, quanto para extrair informações valiosas das análises desses grandes volumes.

4 OS IMPACTOS DO *BIG DATA*: DA VIGILÂNCIA À PRIVACIDADE DOS DADOS

Como se viu, *big data* é muitas vezes saudado por sua capacidade de acrescentar ao conhecimento novas maneiras de enriquecer o entendimento (LAZER et al., 2009; LOHR, 2012). Por outro lado, há que se considerar os riscos e limites dessa capacidade, no que se refere aos usos que se faz dos dados, principalmente os pessoais, que são na maioria das vezes coletados de ambientes das redes sociais digitais, tidos como espaços públicos e sujeitos a diversos usos. Nesse contexto, o *big data* leva às discussões de questões de poder, de vigilância, controle e também privacidade, que serão tratados a seguir, a partir das considerações dos autores como Foucault (1975), Deleuze (1990), Lyon (2007), Bauman (2013), Bruno (2013) e Lemos (2013), entre outros. Busca-se apresentar neste capítulo uma discussão em torno da vigilância exercida em períodos eleitorais e os usos políticos que são feitos das informações e dados coletados em ambientes das mídias sociais.

4.1 SOCIEDADE VIGIADA

Atualmente o indivíduo está sujeito a todos os tipos de vigilância e controle. A vigilância se espalha pelas ruas, pelos prédios, pelos transportes públicos e pelas redes sociais. Busca-se situar nessa perspectiva a conceituação da “sociedade do controle” cunhada por Gilles Deleuze (1990) em comparação à sociedade disciplinar discutida na teoria de Michael Foucault (1975) e a “sociedade líquida” de Zygmunt Bauman (2013), ampliando as discussões com autores que discutem a temática vigilância digital a partir dos “rastros digitais” (BRUNO, 2013; LEMOS, 2013) deixados pelos usuários das mídias digitais.

Como aponta o sociólogo David Lyon na obra de Bauman (2013), a vigilância é uma dimensão-chave do mundo moderno. Não apenas em Londres e Nova York, mas também em Nova Delhi, Xangai e Rio de Janeiro, as câmeras de vídeo são elementos comuns nos lugares públicos. A vigilância é constante e interminável em qualquer ambiente, seja ele físico ou virtual. O fenômeno da vigilância nas sociedades contemporâneas ganhou maior impulso a partir dos atentados de 11 de setembro de 2001 contra as Torres Gêmeas americanas, quando se verificou um crescimento acentuado das câmeras de vigilância instaladas nos Estados Unidos e, posteriormente, no mundo.

Assim, sob tal perspectiva, é possível esboçar a partir de David Lyon, um conceito de *surveillance* (vigilância). Nesse caso, a vigilância toma outra conotação. Ela é a atenção concentrada, sistematizada e rotineira aos dados pessoais cujo objetivo é influenciar, gerenciar, proteger ou dirigir. Em explicação, concentrada, pois seus alvos finais são, via de regra, os indivíduos. Sistematizada, uma vez que essa atenção não é aleatória ou

ocasional: é deliberada e depende de determinados protocolos e técnicas. Rotineira, porque “normalizada”, ou seja, compreendida como parte inescapável do cotidiano em todas as sociedades atuais, uma vez que dependem da associação crescente entre a tecnologia da informação e a administração burocrática. Dessa forma, a *surveillance* é um fenômeno onipresente nas sociedades contemporâneas, podendo ser considerada “[...] one of those major social processes that actually constitute modernity as such”⁵¹ (LYON, 2007, p. 14).

Hoje outros tipos de vigilância têm sido experimentados principalmente na internet, relativos a compras rotineiras e comuns, acesso on-line, participação em mídias sociais. A cada dia, o Google anota as buscas dos usuários, estimulando estratégias de marketing customizadas. A tentativa de saber o máximo possível sobre os usuários tornou-se a batalha fundamental da nossa era entre empresas da internet como Google, Facebook, Apple e Microsoft. Embora o Gmail e o Facebook, sejam ferramentas úteis e gratuitas, também são mecanismos extremamente eficazes e vorazes de extração de dados, nos quais os usuários despejam os detalhes mais íntimos das suas vidas. Por meio de aparelhos de hoje, como *Smartphones*, sabe-se exatamente onde estão seus usuários, para quem ligam, o que leem: graças a sua tecnologia e seu microfone, giroscópio e GPS embutidos, sabe-se que o usuário está caminhando, no carro ou numa festa (PARISER, 2013).

Usando protocolos ultravelozes qualquer site – não só os Googles e Facebooks – pode agora participar da brincadeira. Para os comerciantes do “mercado do comportamento” cada “indicador de clique” que enviamos é uma mercadoria e cada movimento que fazemos com o *mouse* pode ser leiloado em microssegundos a quem fizer a melhor oferta. (PARISER, p. 12-13, 2013)

Para Pariser, quanto mais personalizadas forem suas ofertas de informações, mais anúncios eles conseguirão vender e maior será a chance de que os usuários comprem os produtos oferecidos. A personalização não define apenas aquilo que os usuários da internet compram. Feeds de notícias do Facebook, por exemplo, estão se transformando em sua fonte principal de informações. A popularidade desse site está disparando em todo o mundo: quase meio milhão de pessoas adere a cada dia. Esse debate também recai, sobretudo, em uma nova lógica de controle na internet, que remete às teses sobre vigilância de Foucault (1975, p. 169): “A vigilância torna-se operador econômico decisivo, na medida em que é ao mesmo tempo peça interna no aparelho de produção e engrenagem específica no poder disciplinar”. Com a vigilância hierarquizada, o poder disciplinar torna-se um sistema integrado, ligado à economia e aos fins do dispositivo onde é exercido.

⁵¹ Tradução livre: um dos principais processos sociais que constituem efetivamente modernidade como tal.

Busca-se aqui uma recuperação dessa discussão em *Vigiar e Punir*, em que Michel Foucault (2003) faz da máquina benthamiana a ilustração exemplar da emergência da sociedade disciplinar, dos sonhos e mecanismos de recriação planejada dos indivíduos – simultaneamente muitos e também de cada um. O Panóptico⁵² - olho que tudo vê - é o símbolo mais claro do sonho da disciplinarização radical dos comportamentos, da máxima docilização dos corpos, além da promessa do fim do desvio, do crime, da violência e de comportamentos indesejáveis em geral. Para o filósofo Jeremy Bentham, o Panóptico era uma máquina de produzir a imitação de Deus, uma maneira de recriar os indivíduos, não à sua imagem e semelhança, mas ao seu desejo e necessidade (MILLER, 2000). O Panóptico é um termo utilizado para designar um centro penitenciário ideal desenhado por Bentham em 1785, como mostra a Figura 8.

Da torre central, a visibilidade das celas, para o inspetor, era total. Em compensação, das celas não se via o interior da torre, de onde a vigilância (ou inspeção) era realizada.

Figura 8: Projeto de Bentham



Fonte: Deleuze (1990)

De acordo com o projeto de Bentham, este seria um *design* mais barato que o das prisões de sua época, já que requer menos empregados para fazer a vigilância sobre os presos.

⁵² No século XVIII o filósofo e jurista inglês Jeremy Bentham concebeu pela primeira vez a ideia do panóptico. Para isto Bentham estudou “racionalmente”, em suas próprias palavras, o sistema penitenciário. Criou então um projeto de prisão circular, onde um observador central poderia ver todos os locais onde houvesse presos. Eis o panóptico. Ele também observou que este mesmo projeto de prisão poderia ser utilizado em escolas e no trabalho, como meio de tornar mais eficiente o funcionamento daqueles locais. Foi naquele período da história que, segundo o francês Michel Foucault, iniciou-se um processo de disseminação sistemática de dispositivos disciplinares, a exemplo do panóptico. Um conjunto de dispositivos que permitiria uma vigilância e um controle social cada vez mais eficientes, porém, não necessariamente com os mesmos objetivos “racionalistas” desejados por Bentham e muitos de seus antecessores e contemporâneos.

O Panóptico é tido por muitos como o precursor do *Big Brother* como um exemplo de como o controle quase absoluto poderia existir sem sequer o controlado pudesse perceber em que termos ele ocorria. O termo também é usado pelos teóricos das novas tecnologias, como Pierre Lévy e Dwight Howard Rheingold, para designar o possível controle exercido pelos novos meios de informação sobre seus usuários.

Toma-se como referência esse tipo de propósito da vigilância a partir das discussões estabelecidas por Deleuze (1990). Esse autor diferencia uma sociedade disciplinar de uma sociedade de controle. Segundo ele, para Foucault, as sociedades disciplinares podem ser situadas num período que vai do século XVIII até a Segunda Grande Guerra, sendo que os anos da segunda metade do século XX estariam marcados por seu declínio e pela perspectiva ascensão da sociedade de controle. Foucault (1987) defende que desde o século XII, as disciplinarizações não pararam de ganhar campos cada vez mais vastos e de buscar cobrir todo o corpo social. Elas já existiam nos conventos e nos exércitos, porém, no decorrer dos séculos XVII e XVIII, tornaram-se fórmulas gerais de dominação. Para esse filósofo em qualquer sociedade o corpo está preso no interior de poderes, que lhe impõem limitações, proibições ou obrigações.

O fim do anonimato digital na sociedade da informação pode estar intimamente relacionado com a necessidade de invenção de um novo conceito para Panóptico no século XXI. A capacidade de vigiar os cidadãos no ciberespaço, seja a partir de redes sociais, para criação de propagandas direcionadas ou com a ajuda de *softwares* especializados para controle do Estado, é uma realidade latente no mundo todo.

A ideia da sociedade de controle de Deleuze (1992) parte da constatação da superação da vigilância panóptica de Bentham e da sociedade disciplinar do confinamento discutida por Foucault. Na realidade, os dois regimes convivem hoje, havendo, entretanto, uma inflexão em direção a uma vigilância mais sutil e invisível, mais modular. Para Deleuze (1992), a sociedade do controle era o que Foucault anunciava como o nosso futuro próximo, o que em termos práticos de vigilância significa que as tecnologias não são mais visíveis e imóveis, mas ubíquas, pervasivas, “nas coisas”, difusa, não exigindo do sujeito o confinamento, mas pedindo exatamente o contrário, a mobilidade permitindo um controle dinâmico. Isso nos remete aos exemplos dos monitoramentos dos rastros digitais de usuários, a fim de identificar potencialidades. Assim, a sociedade do controle é caracterizada por Deleuze (1992) pela invisibilidade e pelo nomadismo que se expande nas redes de informação. Segundo ele, as sociedades de controle apontam uma espécie de antiarquitetura. A ausência da casa, do

prédio, do edifício é fruto de um processo em que se caminha para o mundo virtual (DELEUZE, 1992).

Da mesma forma Lemos ao fazer referências à tese de vigilância de Foucault, afirma que a nova vigilância da sociedade de controle está em todos os lugares e, ao mesmo tempo, em lugar nenhum.

Diferente dos “internatos”, os atuais meios de vigilância não se dão mais em espaços fechados, mas nos “controlatos” dos perfis da Internet, nos bancos de dados em redes sociais interconectadas, nos deslocamentos com o telefone celular monitorando o “roaming” do usuário, na localização por GPS, nos rastros deixados pelo uso de cartões eletrônicos, nos smartcards dos transportes públicos, nos sinais emitidos e captados por redes bluetooth, nas etiquetas de radiofrequência que acompanham produtos e compradores... Certamente tudo está menos visível e mais difuso, tornando essa invisibilidade vigilante mais performativa e o controle dos movimentos mais efetivo (LEMOS, 2009, p. 630).

Para esse autor, não se trata mais de fechar e imobilizar para vigiar, mas de deixar fluir o movimento, monitorando, controlando e vigiando pessoas, objetos e informação para prever consequências e exercer o domínio sob as “modulações”. Deleuze sugere ainda que as sociedades disciplinares possuem dois polos, “a assinatura que indica o indivíduo, e o número de matrícula que indica sua posição numa massa”. Nas sociedades de controle, “o essencial não seria mais a assinatura nem um número, mas uma cifra: a cifra é uma senha”. (...) A linguagem digital do controle é feita de cifras, que marcam o acesso ou a recusa a uma informação” (DELEUZE, 1990, p.52). A passagem de um a outro implica que os indivíduos deixam de ser, justamente, indivisíveis, pois passam a sofrer uma espécie de divisão, que resulta do estado de sua senha, de seu código (ora aceito, ora recusado). Dessa maneira, as massas, por sua vez, tornam-se amostras, dados, mercados, que precisam ser rastreados, cartografados e analisados para que padrões de comportamentos repetitivos possam ser percebidos. Mesmo não fazendo referências ao termo da atualidade *big data* em suas discussões, Deleuze já antecipava a finalidade desse conceito numa sociedade informacional (CASTELLS, 1999).

4.1.1 A vigilância distribuída

A vigilância suaviza-se especialmente no reino do consumo, assim percebe Bauman (2013, p.10), que “velhas amarras se afrouxam à medida que fragmentos de dados pessoais obtidos para um objetivo são facilmente usados com outro fim”. Para o autor:

A vigilância se espalha de formas até então inimagináveis, reagindo à liquidez e reproduzindo-a. Sem um contêiner fixo, mas sacudida pelas demandas de

“segurança” e aconselhada pelo marketing insistente das empresas de tecnologia, a segurança se esparrama por toda parte (BAUMAN, 2013, p. 10).

De acordo com o entendimento de Bauman, quanto mais a tecnologia evolui, menos liberdade tem o homem, que passa a ser refém de câmeras de vigilância, de um governo autoritário que intervém em todos os meios de comunicação sem se preocupar com questões éticas e de nós mesmos, nos sucumbindo às estratégias políticas e de marketing e nos tornando reféns do consumismo.

Para esse autor, a vigilância é uma dimensão central da modernidade, mas ele observa que a sociedade não fica parada. Bauman⁵³ (2013) sugere que a modernidade tem se liquidificado de novas e diferentes maneiras (para além do insight de Marx e Engels, na fase inicial da modernidade, de que “tudo é sólido se desmancha no ar”). Bauman cita Lyon (2007) que questiona se a vigilância também se caracteriza assim. A esse respeito, Lyon é categórico em afirmar que diversos teóricos (Deleuze, Haggerty, Ericson, Staples) observam que a vigilância, antes aparentemente sólida e estável, se tornou muito mais móvel e flexível, infiltrando-se e se espalhando em muitas áreas da vida sobre as quais sua influência era apenas marginal.

Bruno traz para discussão a legitimação de uma vigilância moderna, que ela chama de “vigilância distribuída”, como definição do estado geral da vigilância nas sociedades contemporâneas. A autora explica o que seria a noção de distribuição:

A noção de distribuição para busca designar um processo reticular, espreado e diversificado, pleno de ambiguidades, que não se confunde com a ideia de uma vigilância homogênea, sem arestas nem conflitos. Ao contrário, os atuais processos de vigilância só podem se entendidos se levarmos em conta que as ações que os constituem são distribuídas por diferentes setores cujos interesses e perspectivas são múltiplos e não obedecem a nenhum princípio unificado (BRUNO, 2013, p.25).

Assim, trata-se de uma vigilância que tende a se tornar incorporada a diversos dispositivos, serviços e ambientes, mas que se exerce de modo descentralizado, não hierárquico e com uma diversidade de propósitos, funções e significações (BRUNO, 2009). Essa vigilância é encontrada nos mais diferentes setores como nas medidas de segurança e circulação de pessoas, informações e bens, inclusive nas estratégias de consumo e marketing, bem como nas formas de comunicação, entretenimento, sociabilidade e prestação de serviços.

Seria uma vigilância “para todos” com uma variedade de focos possíveis.

⁵³ Na tese de Bauman modernidade líquida é definida como um momento em que a sociabilidade humana experimenta uma transformação em diversos processos. Para esse autor todas as formas sociais se desmancham mais depressa que a velocidade com que se criam novas formas. Elas não podem manter seu molde nem se solidificar em arcabouços de referência para as ações e estratégias de vida dos seres humanos em função da brevidade de sua própria vida útil (BAUMAN, 2013, p 11).

Nota-se que em certos casos ela se exerce misturada a dispositivos que não são prioritariamente voltados para a vigilância, sendo assim uma função potencial ou um efeito secundário de dispositivos que são projetados inicialmente para outras finalidades – comunicação, publicidade, geolocalização etc. Daí deriva uma multiplicidade de objetos atuais ou potenciais da vigilância, que não mais se restringem nem se justificam por grupos suspeitos ou supostamente perigosos, mas que podem ser todos e qualquer um – consumidores, transeuntes, internautas, criminosos, participantes de reality shows etc (BRUNO, 2009, p. 15-16).

Na concepção da autora não se vigiam ou monitoram apenas indivíduos ou grupos, mas informações, transações eletrônicas, comportamentos, hábitos e rastros no ciberespaço, comunicações, fluxos de corpos anônimos no espaço urbano.

Nesse tipo de vigilância distribuída, observam-se diferenças em relação à vigilância hiper-panóptica, a qual supõe sistemas centralizados, hierarquizados, dirigidos a grupos ou indivíduos previamente delimitados cujas identidades supostamente portam uma periculosidade que demanda vigilância e se inscreve num projeto de normalização. Bruno (2009) ressalta que, ainda que aspectos desses sistemas persistam hoje, a vigilância contemporânea se complexificou em relação aos modelos modernos, ganhando novos sentidos, modos de atuação e efeitos.

Assim, a vigilância se distribui por diferentes canais e ambientes de comunicação e informação. “O adjetivo distribuído ressalta, ainda, uma tensão que provém do seu uso na qualificação de formas atuais de comunicação, sobretudo no ciberespaço e suas redes e ambientes recentes em que vigem a interação e a participação dos usuários” (BRUNO, 2009, p 16).

Entende-se, a partir das considerações apresentadas, que, diante de ambientes interativos e colaborativos como as mídias sociais, em tempos de concorrência ou mesmo disputas eleitorais, nenhum espaço fuja dessas práticas e ações de vigilância e monitoramento, com intenções claras e definidas.

4.2 A BUSCA POR PADRÕES DE COMPORTAMENTO

Com a explosão da internet no início dos anos de 1990, muitos foram os sites que começaram a utilizar a declaração de perfil de cada usuário para uma série de operações como ofertas de produtos, notícias, bem como de pares de namoro. No entanto, com o aparecimento das redes sociais na virada do milênio e com o desenvolvimento de novas e potentes tecnologias de coleta de dados, que permite rastrear os perfis de usuários de maneira dinâmica, com o objetivo de acompanhamento de suas atividades e aprendendo sobre seus hábitos e comportamentos. Hoje essas ferramentas trabalham orientadas a dados e a textos e

buscam, por meio de monitoramentos, encontrar padrões de comportamento para diversas finalidades, entre elas até mesmo encontrar padrões no comportamento do voto do eleitor.

Os rastros deixados pelos usuários de redes sociais são classificados e posteriormente servem de indicativos de comportamento e tendências. Nesse sentido cabe explorar os mecanismos que se alimentam dessa exposição, mas que são, segundo Bruno (2013, p. 144), bastante discretos, situando-se sob o fluxo participativo, mas intervindo sobre ele. Aqui, as discussões apresentadas giram em torno dos sistemas de coleta, monitoramento, mineração e classificação do imenso volume de dados pessoais gerados pelas plataformas participativas da web 2.0 e que sustentam os bancos de dados dos projetos de *big data*. A facilidade em estocar e recuperar informações e isso vem se intensificando rapidamente nas últimas décadas (GANDY, 2002). Hoje processos sociais, subjetivos, econômicos, cognitivos tornam-se permeáveis ao monitoramento cotidiano (BRUNO, 2013). Com o desenvolvimento de ferramentas e *softwares* específicos, dados antes custosos e de difícil acesso tornam-se passíveis de coleta regular, automatizada e a distância.

Os métodos de monitoramento vão desde o rastreamento de cliques e a mensuração do tempo dedicado a cada página da internet até a captura automatizada do que o usuário tecla, suas ações quando visita um site. O usuário não tem como fugir desse rastreamento. Cabe ressaltar que essas ações de monitoramento seguem diversos propósitos enfatizando mais uma vez o caráter distribuído dos processos de vigilância. Destaca-se, assim, que a vigilância de dados (CLARKE, 1990) ou de rastros pessoais (BRUNO, 2012) na internet é uma via privilegiada de conhecimento, classificação e intervenção sobre indivíduos e grupos.

Como já discutido, na medida em que as mídias sociais⁵⁴ se tornam cada vez mais usuais no cotidiano das pessoas, opiniões, experiências e expectativas de todo o tipo, sejam em relação às outras pessoas, a produtos, ou à política, são expressas na rede de forma pública.

No entanto, em ambiente acadêmico pouco se discute sobre a importância do monitoramento de redes sociais⁵⁵ sob a ótica de estratégias competitivas. O fenômeno tem sido referenciado na maioria das vezes no âmbito da privacidade e segurança dos dados coletados. Alguns trabalhos (BRUNO, 2013) questionam como organizações (Facebook e Google) utilizam as informações postadas pelos seus usuários.

⁵⁴ Mídias Sociais são ambientes tecnológicos utilizados para o compartilhamento de conteúdo – texto, imagem, vídeo, música – provocando a disseminação de opiniões, ideias e experiências.

⁵⁵ Raquel Recuero pensa as redes sociais na internet como agrupamentos complexos instituídos por interações sociais apoiadas em tecnologias digitais de comunicação (RECUERO, 2009, p. 13).

Lemos (2007) entende o monitoramento como forma de observação para acumular informações visando projeções ou construção de cenários e históricos, ou seja, como uma ação de acompanhamento e avaliação de dados. O monitoramento nada mais é do que aplicar metodologias diversas para acompanhar e compreender o que falam e sobre como falam, seja no ambiente off-line ou on-line. Pode-se também dizer que monitoramento é o ato de acompanhar e avaliar (Dicionário Aurélio).

Castells (2003) explica que as aplicações de *software* podem ser superpostas em camadas e protocolos da internet, tornando possível identificar rotas de comunicação e conteúdo. Com o uso dessas tecnologias, é possível violar a privacidade, e uma vez que se torna possível relacionar indivíduos com processos específicos de comunicação em contextos institucionais específicos, todas as formas tradicionais de controle político e organizacional podem ser lançados sobre o indivíduo em rede. Assim, uma vez que os dados são coletados em forma digital, todos os itens de informação contidos no banco de dados podem ser agregados, desagregados, combinados e identificados de acordo com o objetivo e o poder legal. Por vezes, trata-se simplesmente de fazer perfis agregados, como em pesquisa de mercado, seja para o comércio ou para a política (CASTELLS, 2003, 142).

Dessa forma, o monitoramento de mídias sociais diz respeito a acompanhar e avaliar o que vem sendo dito sobre produtos, serviços e pessoas em sites de redes sociais, com fins de gerenciamento de marca e ideias. As pesquisas de marketing são exemplos de atividades de monitoramento aplicadas há décadas. Trata-se de formas amplamente conhecidas de monitorar os ambientes de marketing. Quando uma empresa está desenvolvendo um novo produto, por exemplo, ela faz pesquisas (como testes com clientes reais, grupos de foco etc.) para entender qual a receptividade daquele produto ou para saber quais são os diferenciais percebidos pelo cliente que diferenciam o produto A do produto concorrente B. Os conhecidos relatórios do Ibope sobre o impacto de programas de TV nos telespectadores, por exemplo, são outras formas de se monitorar a audiência.

Segundo Krishnamurthy (2009), rastrear informações de indivíduos na internet não é algo novo. Já existe há vários anos o rastreamento de padrões de navegação por meio dos *cookies*, que são pequenos arquivos de dados, enviados pelos sites da internet e armazenados nos computadores dos usuários. Esses arquivos podem ser posteriormente consultados ou por outros sites para obter informações sobre sites navegados, pesquisas realizadas, entre outras informações. Por outro lado, Aquino (2012) considera que o uso de cookies ou redes sociais para direcionar propagandas não é um problema, pois muitos usuários estão dispostos a receber anúncios relevantes durante o uso das redes sociais. Os usuários entendem ser esta

uma contrapartida pelo uso gratuito dos serviços destes sites e, como relatado por Tucker (2013), pode ser até interessante para o consumidor ter anúncios feitos sobre medida para a sua necessidade atual. Entretanto, como ressaltado por Krishnamurthy (2009), o uso de cookies de navegação, juntamente com as redes sociais e o poder computacional de soluções de *big data*, pode trazer novas consequências para os usuários, como a geração de informações pessoais identificáveis (IPIs), ou seja, o estabelecimento de vínculos entre informações coletadas pelos cookies (como sites visitados, padrão de navegação e interesses) com as informações da sua rede social (como nome, amigos, endereço).

Mas afinal qual é a função do monitoramento? A principal função do monitoramento é ouvir o que o público fala sobre determinada marca, produto, pessoa, ou mesmo, político e candidato. O monitoramento traça diversos perfis do usuário das redes e busca responder: **(1) onde e quando falou? (2) O quanto se falou? (3) Sobre o que falou? (3) Quem falou?**

De acordo com Tuten (2008), a codificação sobre o que é dito sobre uma marca via análise de conteúdo, através da criação de categorias, é trabalho que vem sendo feito por empresas a partir da combinação da ação de *softwares* específicos para contagem e coleta de menções, com a análise da ocorrência de marcas em conversações em mídia social por equipes especializadas.

Esse monitoramento é feito de maneira informatizada e com auxílio de *softwares* específicos, como mostra o quadro 5. A maioria dessas ferramentas podem ser usadas de forma gratuitas.

Quadro 5: Ferramentas de Monitoramento

TIPO	O QUE FAZ
Alertas do Google	Com os <u>alertas do Google</u> é possível ficar informado de tudo o que é publicado com relação aos concorrentes ou com relação a quem o usa. São definidas palavras-chave que se referem aos objetivos e recebe no seu sempre quando é publicado algum conteúdo relacionado àquela expressão.
SocialMention	Monitora os concorrentes pesquisando como eles estão se saindo. Fornecem dados sobre como a marca ou o produto concorrente estão se apresentando em redes sociais, sites de notícias e blogs especializados.
Topsy	O <u>Topsy</u> é focado no Twitter pode entender como está a influência do concorrente e até mesmo da sua marca, analisando os dados de engajamento em tweets atuais e antigos (desde 2006).
Competitive Research & Keyword Research Gadget	Nesse sistema é possível interligar ao site oficial permitindo a extração de dados sobre as visitas e os visitantes. Com ele pode usar outras ferramentas que ajudem a comparar os dados do site com os dados dos concorrentes, de modo a interpretar como o site oficial pode melhorar.
InfiniGraph	Monitora o compartilhamento nas redes pelas pessoas de marcas, produtos e conteúdo, que segue a empresa ou candidato político.
SEMRush	É uma ferramenta focada em entender o concorrente. Com dados de palavras-chave orgânicas e classificação, palavras-chave do anúncio e o tráfego por tais palavras, terá em mãos os dados que precisa para um acompanhamento do mercado.
Alexa	Esta é uma das ferramentas mais avançadas disponíveis hoje e há um bom tempo classifica a relevância de websites. O <u>Alexa</u> faz tudo quando se trata de acompanhar a concorrência: fornece dados sobre links, tráfego, palavras-chave e mais. Monitora concorrentes internacionais.
Trendrr	Dados estruturados em mais de 50 fontes. Track qualitativo, a influência quantitativa, sentimento, situação e tendências competitivas em tempo real e ao longo do tempo. O Trendrr monitora as mídias sociais e coloca os dados em uma perspectiva numérica.
Social Mention	É uma plataforma para análise e busca em mídias sociais que agrega conteúdos gerados por usuários em um único fluxo de informação. Ele permite que você acompanhe e mensure facilmente o que as pessoas estão dizendo sobre você, sua empresa, um novo produto, ou qualquer assunto em tempo real.
Scup	É uma plataforma desenvolvida pela Direct Labs onde é possível monitorar tecnologias como o Twitter, Facebook, blogs, portais de notícias, Flickr, Youtube e Yahoo Respostas. A ferramenta também permite a publicação de conteúdo, classificar as conversas, adicionar tags e gera gráficos precisos das atividades nas redes sociais.

Fonte: Autora

Para cada projeto de monitoramento é um usado um *software* específico. O grande volume de informações que são gerados nos ambientes da internet requer programas que sejam capazes de fazerem filtragens e capturas das menções e interações, gerando relatórios atualizados. Esse tipo de trabalho ajuda os profissionais de comunicação a entender e orientar muitas das estratégias de comunicação para produtos, serviços, instituições ou mesmo políticos e candidatos em processos eleitorais.

4.3 QUEM VIGIA QUEM NAS REDES SOCIAIS

Como já discutido anteriormente a vigilância que ocorre na internet é marcada pelo monitoramento e registro dos rastros comunicativos deixados pelos internautas em seu uso cotidiano do meio para a classificação e posteriori antecipação de comportamentos, preferências, tendências e interesse. Mas quem é o vigia e quem é vigiado nas redes sociais digitais?

A ideia do Panóptico na estrutura das redes sociais, onde não mais um vigia a todos, mas onde todos vigiam a todos com muito mais artifícios do que os utilizados nos campos periféricos e torres de pensamento de Bentham. O modelo estrutural do Panóptico propõe o exercício do poder pelo olhar, ou seja, pela instituição da visibilidade e da transparência como estratégias de submissão e dissuasão, materializada, na prática a esfera da opinião pública como instantânea preventiva. Assim, “seria o simples fato de que as coisas são sabidas e de que as pessoas serão vistas por um tipo de olhar imediato coletivo e anônimo” (FOCAULT, 1975, p. 216-217).

No ambiente das redes sociais, levando em consideração as suas características, parte-se do princípio que a vigilância é esperada pelos internautas. Uma vez que a produção de conteúdos é pensada em atingir determinado público na rede e obter determinada “audiência”.

A respeito disso Bruno afirma que:

Hoje, a crescente exposição da vida íntima e privada nos diversos meios de comunicação e a forte presença da imagem na relação que os indivíduos estabelecem com o mundo, com o outro e consigo apontam para uma subjetividade radicalmente próxima do olhar do outro e inserida num regime de visibilidade onde as tecnologias de informação e de comunicação têm um papel fundamental (BRUNO, 2005, p.58).

Dessa forma, existe uma racionalidade em se expor que talvez não calcule certas consequências que estão por trás desse monitoramento e vigilância que se apropria de conteúdos para estabelecer características de perfis. Existe, por parte do internauta criador de conteúdos e fornecedor de informações, um desejo em ser vigiado. Quando por exemplo ele pensa em manifestar a sua opinião sobre política ou algum candidato em um período eleitoral, ele tem a intenção de ter a sua mensagem visualizada, muitas vezes curtida e compartilhada, tanto por simpatizantes da mesma ideia ou não. Da mesma forma ele busca de alguma forma chamar a atenção desse candidato ou político, assim como de outros para que saibam o que ele pensa deles. Sobre a necessidade de visibilidade Dal Bello, diz que:

O que muda, em essência, não é a paridade entre visibilidade e vigilância, mas a predisposição pessoal à exibição pública e sua crescente naturalização. A Geração Y não parece se incomodar com as possibilidades de

rastreamento, controle e indexação a que está sujeita quando se inscreve nas redes (DAL BELLO, 2011, p. 147).

Para esse autor a explosão demográfica das redes sociais digitais, que possibilitam a manifestação subjetiva e a promoção do “eu” para uma audiência cativa, formada por amigos ou seguidores, assinala o quanto tais plataformas parecem corresponder à necessidade de ser reconhecidamente alguém. Nesse sentido, segundo Dal Bello (2011) confundir-se tão intimamente com o imaginário do espetáculo, a visibilidade facultada pelas redes torna-se imprescindível e desejável – ainda que traga consigo, pois lhe é intrínseca!, a faceta da vigilância.

E é nesse ponto que reside o valor mercadológico de uma rede social: ela congrega indivíduos que a utilizam como base de apresentação, comunicação e ambiente de relacionamento; classifica-os em grupos de perfis combinando variáveis diversas; rastreia deslocamentos coletivos ou individuais de interesse; indexa, a qualquer tempo, qualquer um para entregar-lhe mensagens em geral publicitárias, particularmente significativas.

Assim, a condição do processo panópticos, em que poucos vêem muitos, podem também surgir dentro das redes sociais digitais, com campos de visibilidade efetivamente assimétricos e programados para atender a interesses específicos. É assim que se configuram os monitoramentos praticados por exemplo por candidatos e políticos em campanhas eleitorais.

A prática de monitoramento com finalidades políticas tem sido vista no Brasil desde 2010 e vem se intensificando a cada eleição. Para alguns especialistas, a preocupação com o que dizem os eleitores sobre políticos e candidatos é uma consequência do efeito Obama, que em 2008 usou efetivamente as redes sociais em sua campanha à presidência dos Estados Unidos⁵⁶. Assim, no Brasil se já era uma prática forte nas eleições anteriores, nas eleições de 2010 uma força profissional já especialista na área, dispendo de ferramentas plenas de monitoramento, pôde orientar estrategistas políticos sobre as flutuações nas opiniões do eleitorado, militância e detratores.

Nas eleições de 2012, os profissionais formados nos últimos anos geraram ainda mais cases, sobretudo nas eleições para os cargos executivos municipais. Pesquisa⁵⁷ lançada em julho de 2012 mostrou que 75% dos deputados federais estavam presentes de forma relevante ou influente na internet e mídias sociais, número ainda maior entre os senadores: 84%. De

⁵⁶ Eleições 2008. O Brasil e o efeito Obama, por Gustavo Fleury.

⁵⁷ Dados retirados da pesquisa Político 2.0, realizada pela Medialogue. Disponível em: www.medialogue.com.br. Acesso em: 12 de maio de 2015.

2010 até 2013 políticos melhoraram suas redes sociais e intensificaram ainda mais os seus monitoramentos, como aponta a Pesquisa Política 2.0 realizada pela Medialogue⁵⁸.

Nas eleições de 2014 os processos de monitoramentos foram mais intensificados, principalmente a partir das manifestações de 2013, que ganharam as ruas, a partir das iniciativas das redes sociais digitais, que programaram e organizaram protestos em diversas cidades brasileiras contra principalmente a corrupção política e insatisfações com as administrações públicas.

Assim, nos últimos anos as redes sociais se tornaram um espaço de discussão para problemas e situações de questões públicas e o número de menções em relação a isso só aumenta – ainda mais quando vivemos situações de movimentação política como durante o período eleitoral.

No caso brasileiro, sem detectar as manifestações combinadas pelas redes sociais e que tiveram como alvo o Palácio do Planalto, a Agência Brasileira de Inteligência (Abin) montou as pressas uma operação para monitorar a internet⁵⁹. Segundo a reportagem do jornal O Estadão com a eclosão da crise, o potencial das manifestações passou a ser medida e analisada diariamente pelo Mosaico, sistema online de acompanhamento de cerca de 700 temas definidos pelo ministro-chefe do GSI, general José Elito. Os relatórios elaborados a partir dos monitoramentos das redes sociais, mostram o roteiro e o tamanho dos protestos, infiltrações de grupos políticos e até supostos financiamentos dos eventos. Esse tipo de monitoramento teve como objetivo por parte do governo de manter o conhecimento e o controle das manifestações e possíveis proliferações dos atos futuros, principalmente na disputa eleitoral de 2014.

Assim, a vigilância, controle e monitoramento das redes sociais digitais se dá por governos, instituições partidárias, políticos, candidatos, empresas de comunicação, empresas de sites de redes sociais digitais e também pelos eleitores/cidadãos seja em períodos eleitorais ou mesma fora deles, em momentos de intenso debates sobre as políticas públicas.

4.4 DADOS COLETADOS PELAS MÍDIAS E REDES SOCIAIS

Como se viu anteriormente, hoje as empresas de mídias sociais vivem seus negócios a partir dos dados e informações dos usuários de seus sites. Este item do trabalho se propõe a demonstrar de forma instrumental que tipos de dados dos usuários são coletados pelas mídias

⁵⁸ Disponível em: <http://www.abradi.com.br/wp-content/uploads/2015/05/Mdialogue-Mapa-dos-Poder-Politico-2014.pdf>. Acesso em: 12 de maio de 2015.

⁵⁹ Disponível em: <http://sao-paulo.estadao.com.br/noticias/geral,abin-monta-rede-para-monitorar-internet,1044500>. Acesso em: 19 junh. 2014.

sociais a partir dos Termos de Usos e Contrato dos serviços e que usos são feitos a partir dessa coleta. Destacam-se aqui as principais mídias sociais como *Facebook*, *Google*, *Instagram*, *Waze*.

O Facebook é uma rede social lançada em 2004, fundado por Mark Zuckerberg, Eduardo Saverin, Andrew McCollum, Dustin Moskovitz e Chris Hughes, na época estudantes da Universidade Harvard. Este termo é composto por *face* (que significa cara em português) e *book* (que significa livro), o que indica que a tradução literal de facebook pode ser "livro de caras". Inicialmente, a adesão ao facebook era restrita apenas para estudantes da Universidade Harvard, e logo foi a muitas universidades individuais. Hoje o uso da rede é gratuito para os usuários e gera receita proveniente de publicidade, incluindo *banners* e grupos patrocinados. Os usuários criam perfis que contêm fotos e listas de interesses pessoais, trocando mensagens privadas e públicas entre si e participantes de grupos de amigos.

A visualização de dados detalhados dos membros é restrita para membros de uma mesma rede ou amigos confirmados, ou pode ser livre para qualquer um. O site de relacionamento possui várias ferramentas, como o mural, que é um espaço na página de perfil do usuário que permite aos amigos postar mensagens para ele ver. Ele é visível para qualquer pessoa com permissão para ver o perfil completo, e posts diferentes no mural aparecem separados no "Feed de Notícias". O "Face", como conhecido popularmente entre os seus usuários, possui também aplicativos, com os mais diversos assuntos, e eventos, pelos quais a pessoa pode convidar todos seus amigos para um determinado evento. Existem versões diferentes do Facebook para telefones celulares e *smartphones*, que facilitam a visualização e acessibilidade dos usuários. O Face é considerado uma das maiores redes de relacionamento, contabilizando 1,4 bilhão de usuários no mundo⁶⁰ em 2015. E esse número se altera a cada segundo. Como mostra o Quadro 6 em que o Facebook⁶¹ utiliza diversos tipos de dados pessoais.

⁶⁰ Informações retiradas do portal da UOL. Disponível em: <http://blogs.ne10.uol.com.br/mundobit/2015/01/29/facebook-tem-14-bilhao-de-usuarios-no-mundo-e-supera-receitas/>. Acesso em: 20 de abril de 2015.

⁶¹ Informações retiradas dos Termos de Uso do site do Facebook. Disponível em: < <https://www.facebook.com/about/privacy>>. Acesso em: 20 de abril de 2015.

Quadro 6: Dados coletados pelo Facebook

Tipos	Dados
Dados de cadastro	nome, e-mail, data de nascimento, sexo, número de telefone;
Informações públicas por padrão	nome, fotos do perfil, fotos de capa, gênero, número de identificação do usuário, redes;
Dados que o usuário define como público	postagens de texto, vídeo, fotos, comentários em posts de amigos, atualização de status e ações como adicionar um amigo, curtir uma página, indicar que está em um relacionamento, adicionar um lugar;
Dados que amigos e outros compartilham sobre o usuário	informações de contato, fotos, atualização de status, localização;
Dados do registro	hora, data e local que fotos e vídeos foram postados, endereço IP ou número do telefone celular, sistema operacional, tipo de dispositivo, navegador. Também dados de quando o usuário visita jogo, aplicativo ou site que usa a plataforma da rede social;
Outros dados	quando o usuário entra no Facebook, olha a linha do tempo de outra pessoa, envia e recebe mensagens, procura um amigo ou página, clica, visualiza ou interage com as coisas, usa um aplicativo da empresa, faz compras no Facebook.

Fonte: Autora

Segundo o site da rede, o Facebook usa os dados pessoais coletados para fornecer serviços como sugerir amigos; sugerir pessoas para marcar nas fotos; informar sobre eventos e promoções próximas à sua localização, escolher histórias para o *feed* de notícias.

Assim, da mesma forma o Google⁶² faz coleta de dados de seus usuários como demonstra o Quadro 7.

Quadro 7: Dados coletados pelo Google

Tipos	Dados
Dados fornecidos pelo usuário para criar uma conta	nome, endereço de e-mail, número de telefone ou cartão de crédito;
Dados de armazenamento local	informações inclusive pessoais – armazenadas nos dispositivos;
Dados de aplicativos instalados	como tipo de sistema operacional e número da versão do app;
Informações do dispositivo	modelo de hardware, versão do sistema operacional número de telefone;
Informações a partir do uso dos serviços	preferências, mensagens do Gmail, perfil do Google+, fotos, vídeos, histórico de navegação, pesquisas de mapas, documentos ou outros conteúdos hospedados no Google;
Dados de registros	registros de telefonia (número de telefone, número de quem chama, horário de data das chamadas etc), endereço IP, cookies, tipo de navegador, configurações do hardware etc;
Dados de localização	como sinais de GPS

⁶² Informações retiradas dos Termos de Uso do site Google. Disponível em: < <https://www.google.com.br/intl/pt-BR/policies/terms/regional.html> >

Segundo as políticas de coleta e usos dos dados dos usuários, o Google faz os seguintes usos dos dados coletados dos usuários: para fornecer, manter, proteger e melhorar os serviços, desenvolver novos e proteger o Google e os usuários; para fornecer um conteúdo mais específico, como resultados mais relevantes de pesquisas e anúncios. O fornecimento e melhoria dos serviços pelo Google incluem também: combinar informações entre os serviços do Google (exemplo: resultados de buscas que mostrem as postagens de amigos no Google+; conectar o usuário a outras pessoas (ex: sugestões de amigos no Google+)); número de telefone para enviar um SMS, caso o usuário esqueça senha da conta; completar automaticamente o endereço do de alguém com quem a pessoa já se comunicou.

O *Instagram*⁶³ é uma rede social on-line de compartilhamento de foto e vídeo que permite aos seus usuários tirar fotos e vídeos, aplicar filtros digitais e compartilhá-los em uma variedade de serviços de redes sociais como Facebook, Twitter, Tumblr e Flickr. Uma característica distintiva é que ela limita as fotos para uma forma quadrada, semelhante ao Kodak Instamatic e de câmeras Polaroid, em contraste com a relação a proporção de tela de 16:9 agora tipicamente usado por câmeras de dispositivos móveis. Os usuários também são capazes de gravar e compartilhar vídeos curtos com duração de até 15 segundos. O Instagram foi criado por Kevin Systrom e Mike Krieger e lançado em outubro de 2010. O serviço rapidamente ganhou popularidade, com mais de 100 milhões de usuários ativos em abril de 2012. Em dezembro de 2014, a rede social chegou a atingir 300 milhões de usuários⁶⁴. O Quadro 8 demonstra que tipos de dados são coletados pelo *Instagram*.

Quadro 8: Dados coletados pelo Instagram

Tipos	Dados
Dados fornecidos pelo usuário	nome de usuário, senha, endereço de , informações do perfil (nome, foto e telefone), conteúdo postado (foto, vídeos, comentários);
Lista de contatos do celular	se o usuário optar por encontrar amigos por ela, e informações que redes sociais transmitirem ao <i>Instagram</i> , caso faça a opção por encontrar amigos por uma rede
Informações analíticas	para medir tráfego, como páginas da web visitadas, além de cookies;
Metadados	são dados técnicos associados com o conteúdo do usuário, como hashtags, comentários e geolocalização em fotos públicas;
Identificadores de dispositivos	que são arquivos de dados que identificam qual o hardware ou sistema operacional usado;
Dados de registro	endereço IP, tipo de navegador, URLs, páginas vistas.

Fonte: Autora

⁶³ Informações retiradas do site do Instagram. Disponível em: < <https://instagram.com/#>>. Acesso em: 03 de abril de 2015.

⁶⁴ Informações retiradas do site O Povo on-line. Disponível em: < <http://www.opovo.com.br/app/maisnoticias/tecnologia/2014/12/10/noticiastecnologia.3361042/instagram-ultrapassa-twitter-em-numero-de-usuarios.shtml>>. Acesso em 03 de Abril de 2015.

O *Instagram* usa os dados coletados de usuários com os seguintes propósitos: fornecer, testar, melhorar e monitorar o serviço. Isso inclui algumas ações como lembrar-se de informações para que o usuário não precise digitar de novo, diagnosticar e consertar problemas técnicos, desenvolver novas funções; fornecer conteúdo personalizado, como publicidade; monitorar métricas, como número de visitantes e tráfego; usar os metadados para criar concursos, ofertas especiais ou outros eventos e atividades no serviço.

O Waze é uma aplicação para smartphones ou dispositivos móveis similares, baseada na navegação por satélite (ex.: GPS) e que fornece informações em tempo real e informações de usuários e detalhes sobre rotas, dependendo da localização do dispositivo portátil na rede. Foi desenvolvida pela start-up Waze Mobile de Israel, empresa que foi adquirida pela Google em 2013. No Quadro 9 são apresentados os tipos de dados coletados pelo aplicativo.

Quadro 9: Dados coletado pelo Waze

Tipos	Dados
Localização e rotas	na forma de sinais de GPS e dados enviados pelo celular;
Metadados	frequência e duração do uso do app, páginas visitadas via app, informações lida, anúncios clicados ou vistos, comunicação com outros usuários ou terceiros, endereço IP;
Informações de redes sociais integradas,	Como Facebook
Informações para criar uma conta	Como nome e sexo
Dados fornecidos pelo usuário	nome de usuário, locais salvos como “favoritos” ou “casa”, destinos, mensagens do chat, termos buscados números de telefones da lista de contatos para encontrar amigos, seu número de telefone para verificar a conta.

Fonte: Autora

O aplicativo Waze usa os dados coletados dos usuários com as seguintes finalidades: para fornecer o serviço de navegação e rotas, encontrar amigos, melhorar o serviço e mapas, lidar com erros e problemas técnicos; para enviar atualizações, anúncios e informações adicionais do serviço, mostrar anúncios no aplicativo de acordo com a localização da pessoa, e enviar publicidade ao e-mail cadastrado, caso o usuário aceite receber; para agregar dados anônimos para estatísticas e melhora do serviço, conduzir pesquisas, prevenir fraudes, tomar ações em casos de disputa e processos legais, e contatar o usuário quando for necessário.

4.5 AS IMPLICAÇÕES DO BIG DATA NA COLETA DE DADOS PESSOAIS

A intensa coleta de dados e informações pessoais possibilitada pelos rastros digitais (BRUNO, 2013; LEMOS, 2013) e compartilhamento destas com outras organizações sem o consentimento dos usuários, podem representar inúmeros riscos à privacidade e devem ser discutidos de forma mais crítica sob diversos aspectos.

Um fator preocupante é que essa grande massa de dados vem servindo de fonte para coleta de informações de usuários, sendo inclusive utilizada e analisada por governos, instituições, empresas ou mesmo por candidatos em disputas eleitorais. Empresas de mídias sociais como o Facebook e o Google conseguem reunir grandes quantidades de dados através da oferta de serviços gratuitos de *Cloud Computing*⁶⁵, ou até mesmo por aqueles capazes de pagar por informações que irão beneficiá-los de alguma forma, tanto socialmente, politicamente e financeiramente.

Tene (2012) alerta que, apesar de existirem políticas de privacidades nas redes sociais, inclusive apresentadas para os usuários no momento de inscrição nos serviços como Facebook ou LinkedIn, estas muitas vezes servem apenas como isenção de responsabilidade para as empresas, em vez de balizadores das garantias de privacidade para os consumidores.

Quando Eric Schmidt, CEO do Google, foi questionado durante uma entrevista à CNBC em 2009 sobre as preocupações em relação à retenção de dados dos usuários praticada por sua empresa, sua resposta foi: *“Se você tiver alguma coisa que não quer que ninguém saiba, talvez não a devesse estar fazendo, para começo de conversa”*. Da mesma forma, o fundador do Facebook, Mark Zuckerberg, afirmou em uma entrevista em 2010 que *“todos já se sentem à vontade não só para compartilhar mais informações de diferentes tipos, mas também de modo mais aberto e com mais pessoas”*. Para ele a privacidade na era digital não é mais uma “norma social”.

Assim, a prática do *big data* tornou-se mais comum do que deveria por muitas empresas e pessoas com finalidades diversas. A respeito disso, Schönberger-Mayer e Cukier afirmam que:

Com o *big data* prometendo valiosas ideias para aqueles que os analisam, todos os sinais parecem apontar para uma nova onda de coleta, armazenamento e reutilização de nossos dados pessoais. O tamanho e escala dos bancos de dados aumentarão à medida que os custos de armazenamento continuarem a diminuir e as ferramentas analíticas se tornarem ainda mais potentes (SCHÖNBERGER- MAYER; CUKIER, 2013, p. 105).

⁶⁵ Também conhecida no Brasil como *computação nas nuvens* ou *computação em nuvem*, a *cloud computing* se refere, essencialmente, à noção de utilizarmos, em qualquer lugar e independente de plataforma, as mais variadas aplicações por meio da internet com a mesma facilidade de tê-las instaladas em computadores locais.

Esses autores ainda levantam alguns questionamentos sobre isso, dizendo que se a era da internet pôs em risco a privacidade, será que o *big data* ameaça ainda mais? Será esse o lado negro do *big data*?

Assim, a partir desses questionamentos, é possível empreender uma discussão sobre que riscos as previsões de *big data* impõem não somente em relação à privacidade, mas ameaças principalmente quando se utilizam essas previsões que tenham por finalidade punições, exemplos desses usados para detectar possíveis terroristas. Segundo Schönberger-Mayer e Cukier (2013, p.106) “existe outro perigo que é o risco de sermos vítimas da ditadura dos dados, na qual adoramos as informações e os resultados de nossas análises e acabamos usando-os de forma equivocada”. Para eles, o *big data*, se usado sem sabedoria, torna-se instrumento de poderosos, que podem transformá-lo numa fonte de repressão, seja ao frustrar clientes e funcionários, ou, pior, ao atacar cidadãos.

Ainda, os autores são categóricos em afirmar que boa parte dos dados gerados hoje inclui informações pessoais. Nesse sentido, as empresas, governos e pessoas têm vários incentivos para captar ainda mais dados, mantê-los por mais tempo e reutilizá-los com mais frequência. “Os dados podem não parecer informações pessoais explícitas, mas, com os processos de análise, podem facilmente dizer a quem se referem ou facilitar a dedução de detalhes íntimos da vida de uma pessoa” (MAYER-SCHÖNBERGER; CUKIER, 2013. p. 106).

Assim, a questão mais importante a ser debatida não é se o *big data* aumenta os riscos para a privacidade, mas que tipos de riscos os indivíduos correm de ter seus dados utilizados para outras finalidades além das apresentadas inicialmente na coleta. Segundo Schönberger-Mayer e Cukier (2013), com o *big data*, o valor da informação não está mais somente no propósito primário, ele está agora nos usos secundários. E dessa forma, essa mudança prejudica o papel central de cada pessoa nas atuais leis de privacidade. Hoje, dizem em que momento a informação está sendo coletada e com qual objetivo.

Assim, o indivíduo tem a oportunidade de concordar. Porém, “saber e consentir” não é a única maneira legal de coletar e processar dados pessoais, é preciso discutir outras formas de proteger os indivíduos de que seus dados possam ser reutilizáveis com outras finalidades, que não sejam autorizadas.

Nesse caso, é preciso rever as leis em vigência que tratam da coleta, armazenamento e usos dos dados pessoais. Cabe neste trabalho uma abordagem ao direito de privacidade dos dados dos indivíduos.

4.5.1 Direito à privacidade dos dados

Como viu-se nas discussões os meios para adquirir e controlar consumidores são múltiplos. Laymert Santos (2000), analisando o emprego e a difusão dos dados individuais, demonstra que a integração de bancos de dados permite a leitura de individualidades, com o estabelecimento de perfis que antecipam desejos e necessidades. Porém, é preciso entender que toda individualidade tem seu direito à privacidade de dados e informações. Vale aqui uma discussão sobre conceitos, direitos e garantias.

Privacidade vem do inglês *privacy* e se constitui no direito à reserva de informações pessoais e da própria vida privada. O jurista norte-americano Louis Brandeis juntamente com Samuel Warren foram os primeiros a formular o conceito de direito à privacidade. O conceito foi inspirado na leitura da obra do filósofo Ralph Waldo Emerson, que propunha a solidão como critério e fonte de liberdade.

O artigo 12 da Declaração Universal dos Direitos Humanos adotada pela Assembleia Geral das Nações Unidas diz que o direito à vida privada é um direito humano:

"Ninguém será objeto de ingerências arbitrárias em sua vida privada, sua família, seu domicílio ou sua correspondência, nem de ataques a sua honra ou a sua reputação. Toda pessoa tem direito à proteção da lei contra tais ingerências ou ataques."

Já o artigo 17 do Pacto Internacional dos Direitos Civis e Políticos adotado pela Assembleia Geral das Nações Unidas consagra, a esse respeito, o seguinte:

1. *"Ninguém será objeto de ingerências arbitrárias ou ilegais em sua vida privada, sua família, seu domicílio ou sua correspondência, nem de ataques ilegais a sua honra e reputação."*
2. *Toda pessoa tem direito à proteção da lei contra essas ingerências ou esses ataques."*

O professor de Direito da Universidade Federal de Minas Gerais, Túlio Vianna (2007), divide o direito à privacidade em três outros direitos que, em conjunto, caracterizam a privacidade, que nos leva a uma reflexão nos dias de hoje em que informações e dados são constantemente objetos de coleta e análise. São eles:

1. Direito de não ser monitorado, entendido como direito de não ser visto, ouvido etc.
2. Direito de não ser registrado, entendido como direito de não ter imagens gravadas, conversas gravadas etc.
3. Direito de não ser reconhecido, entendido como direito de não ter imagens e conversas anteriormente gravadas publicadas na Internet em outros meios de comunicação.

Vianna (2007) observa que o direito à privacidade é concebido como uma tríade de direitos transcende, pois, nas sociedades informacionais, os limites de mero direito de interesse privado para se tornar um dos fundamentos do Estado democrático de direito⁶⁶.

Da mesma forma, Rainer Kuhlen concebe o conceito de “privacidade” (*Privatheit*) não apenas como proteção de dados ou como o direito de ser deixado em paz, mas também como “autonomia informacional”, que é a capacidade de escolher e utilizar o conhecimento e a informação autonomamente, em um ambiente eletrônico, e de determinar quais atributos de si serão usados por outros. A discussão em torno disso tem ganhado muitos espaços, principalmente porque a promessa do *big data* é analisar dados de todos os tipos e fontes e deles tirar relações e projeções para diversas finalidades.

Mas, em termos legislativos, qual é o cenário encontrado sobre o uso de dados pessoais, principalmente os que vêm de redes sociais digitais? No Brasil desde janeiro de 2015, o Ministério da Justiça lançou debate público on-line⁶⁷ (na Internet) em torno da legislação de proteção a dados pessoais. O anteprojeto de lei deverá ser apresentado ao Congresso Nacional e tem como proposta coibir abusos daqueles que utilizam dados pessoais. Os princípios básicos desse projeto preveem o uso razoável dos dados, limitado, relevante e específico dos dados de cada cidadão.

A discussão em torno desse tema segue a linha das iniciativas tomadas por outros países de avanços legislativos para a proteção da privacidade dos cidadãos na era digital. Em janeiro de 2015, o presidente dos Estados Unidos, Barack Obama, propôs três novos projetos de lei que visam proteger os norte-americanos e o rastro de dados que deixam em smartphones, computadores e outros dispositivos. Na Europa, a Diretiva Europeia de Proteção de Dados (EC 95/46) está em revisão desde 2012. As principais propostas de alterações versam sobre transferência de dados para outros países, desenho de perfis e jurisdição.

⁶⁶ O estado de direito é uma situação jurídica, ou um sistema institucional, no qual cada um é submetido ao respeito do direito, do simples indivíduo até a potência pública. O estado de direito é assim ligado ao respeito da hierarquia das normas, da separação dos poderes e dos direitos fundamentais. Em outras palavras, o estado de direito é aquele no qual os mandatários políticos (na democracia: os eleitos) são submissos às leis promulgadas.

⁶⁷ Foram lançados dois portais na Internet para captar sugestões da sociedade. Twitter e Facebook serão canais auxiliares das discussões nas redes sociais. O objetivo da consulta pública é agregar contribuições de forma democrática e participativa. O debate busca promover a participação da sociedade brasileira na elaboração do anteprojeto de lei para proteção de dados pessoais, por meio da formulação de comentários e sugestões sobre o texto proposto. Em entrevistas a diversos meios de comunicação, o Ministério da Justiça considera fundamental ter um marco legal de proteção de dados no Brasil baseado no consentimento e no uso legítimo desses dados, ferramentas de exercício de direitos e padrões mínimos de segurança e privacidade para o cidadão. Atualmente, mais de 100 países já possuem leis de proteção de dados pessoais. (Wikipedia) Disponível em: < https://pt.wikipedia.org/wiki/Estado_de_direito>

Hoje no Brasil existem outras legislações de regulamentação de informações, mas que para muitos especialistas no tema são dispersas e diferentes que trazem alguma garantia à privacidade, sem abranger o tema por completo. Entre estas, destacam-se a Lei de Cadastro Positivo, a Lei de Acesso à Informação, o Marco Civil da Internet, além de alguns dispositivos constitucionais genéricos, como os artigos 5.º, 10.º e 12.º da Constituição. Mas o que regulamenta cada uma dessas leis?

A Lei de Cadastro Positivo nº 12.414 de 9 de junho de 2011 regulamenta a formação e consulta a bancos de dados pessoais ou jurídicos para formação de conjuntos de dados financeiros ou históricos de crédito. Nela já há a garantia à privacidade no tratamento desses dados. Além disso, ela regula a objetividade, veracidade, clareza e facilidade de compreensão para a coleta dos dados que são utilizados para avaliar a situação econômica do titular. A lei garante também o acesso a todos os dados armazenados, além da responsabilidade sobre a atualização e correção de informações obtidas.

A Lei prevê ainda no Parágrafo 3º que ficam proibidas as anotações de:

I - informações excessivas, assim consideradas aquelas que não estiverem vinculadas à análise de risco de crédito ao consumidor; e

II - informações sensíveis, assim consideradas aquelas pertinentes à origem social e étnica, à saúde, à informação genética, à orientação sexual e às convicções políticas, religiosas e filosóficas.

O artigo 5º também determina os seguintes direitos ao cidadão cadastrado:

I - obter o cancelamento do cadastro quando solicitado;

II - acessar gratuitamente as informações sobre ele existentes no banco de dados, inclusive o seu histórico, cabendo ao gestor manter sistemas seguros, por telefone ou por meio eletrônico, de consulta para informar as informações de adimplemento;

III - solicitar impugnação de qualquer informação sobre ele erroneamente anotada em banco de dados e ter, em até 7 (sete) dias, sua correção ou cancelamento e comunicação aos bancos de dados com os quais ele compartilhou a informação;

IV - conhecer os principais elementos e critérios considerados para a análise de risco, resguardado o segredo empresarial;

V - ser informado previamente sobre o armazenamento, a identidade do gestor do banco de dados, o objetivo do tratamento dos dados pessoais e os destinatários dos dados em caso de compartilhamento;

VI - solicitar ao consulente a revisão de decisão realizada exclusivamente por meios automatizados; e

VII - ter os seus dados pessoais utilizados somente de acordo com a finalidade para a qual eles foram coletados.

Já Lei de Acesso à Informação nº 12.527 sancionada pela presidência da República em 18 de Novembro de 2011, em seu artigo 31, também regula como será feito o tratamento das informações pessoais. Nela, são mencionados a necessidade de transparência, respeito à intimidade, vida privada, honra e imagem das pessoas, bem como às liberdades e garantias individuais. Outro ponto importante é a garantia da restrição de acesso por parte de agentes públicos sem autorização para tanto, a não ser que haja: (1) necessidade de prevenção ou diagnóstico médico; (2) estatísticas e pesquisas de interesse público, conforme disposto na lei; (3) cumprimento de ordem judicial; (4) proteção dos direitos humanos; ou (5) proteção do interesse público preponderante. Além disso, há a regulamentação sobre o fornecimento para terceiros mediante assinatura de termo de responsabilidade que deixe claros os objetivos do requerente, bem como as suas obrigações conforme a lei.

Por fim, temos o Marco Civil da Internet⁶⁸, que está em vigor desde junho de 2014. Essa Lei traz garantias gerais para a privacidade. Exemplo é o que diz no artigo 7: os contratos de prestação de serviço na Internet devem ser bem claros com relação à proteção, coleta, armazenamento e tratamento dos dados. Na prática, significa que, ao utilizar um aplicativo ou um serviço on-line (que pode ser uma rede social), o usuário deverá ser informado expressamente sobre a coleta, e consentir com o uso dos seus dados para posterior análise. Além disso, há ênfase na questão da proibição da cessão de informações a terceiros sem autorização prévia. O consentimento expresso é reforçado e exige destaque do assunto nas cláusulas contratuais. A exclusão dos dados dos usuários deve ser garantida àqueles que decidirem apagar seus dados após o término de relação entre as partes. Entre outras garantias, a lei estabelece que o acesso a dados armazenados, de fluxo de informações e comunicações privadas somente poderá ocorrer mediante autorização judicial.

A respeito das informações dos cidadãos, a Lei prevê no artigo 31 que o tratamento das informações pessoais deve ser feito de forma transparente e com respeito à intimidade, vida privada, honra e imagem das pessoas, bem como às liberdades e garantias individuais.

O Marco Civil traz também determinações importantes para os provedores estrangeiros, exigindo que todos os serviços e aplicativos acessados por usuários brasileiros estejam em conformidade com a lei (*compliance*). Foi prevista uma série de sanções para eventual descumprimento, que pode levar até à suspensão das atividades no Brasil.

⁶⁸ Informações sobre a Lei disponível na página do Planalto. http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm

Mas essa preocupação deve ser levada em conta também por empresas que fazem coleta, guarda e análise de dados pessoais. A punição para o vazamento de informações pessoais encontra amparo não apenas no Marco Civil, mas também no Código Civil, na forma de indenização (art. 927).

A partir dessas discussões, é preciso pensar com mais cuidado essa questão dos usos de dados pessoais. Segundo Schönberger-Mayer e Cukier (2013), na era do *big data*, as três principais estratégias usadas para garantir a privacidade que seriam o consentimento individual, opção de exclusão e anonimização, perderam de certa forma a eficiência. E nesse sentido, até o momento nenhuma lei pode garantir de fato que a privacidade não venha a ser violada, já que em todos os lugares que o indivíduo está ou vai sempre estará deixando seus rastros e que serão passíveis de captura e análises.

PARTE II – O CAMPO DO OBJETO DE PESQUISA

5 CAMPANHAS ELEITORAIS: CONTEXTUALIZAÇÃO E CARACTERIZAÇÃO

Campanhas eleitorais são o campo em que se insere o objeto *big data* nesta investigação. Neste capítulo, contextualiza-se de forma genérica as temáticas que envolvem o processo de uma campanha eleitoral. Assim, buscando caracterizar as campanhas e relacionar os ambientes comunicacionais, como espaços importantes e determinantes dos processos políticos na contemporaneidade. Dessa forma, fazendo referência às mudanças da comunicação de massa para a comunicação segmentada, muito usada por campanhas eleitorais nos dias de hoje.

Além disso, são apresentadas discussões em torno do marketing eleitoral e sua dinâmica a partir das pesquisas de opinião pública presentes nas mídias sociais. Ressalta-se também a importância da internet na construção das campanhas eleitorais como espaços de divulgação, mas também de coleta de dados e informações de eleitores, que contribuem para a condução das ações de marketing. São apresentadas discussões teóricas sobre campanhas centradas em dados, a partir dos conceitos de *microtargeting*⁶⁹, que é colocado como uma nova maneira de fazer campanhas eleitorais na atualidade e que a partir das análises de bancos de dados de eleitores, buscam entender o comportamento do eleitor para conduzir as ações de comunicação e marketing eleitoral. E por fim, ilustra-se esta pesquisa com a campanha eleitoral americana de Barack Obama de 2012, que é considerada um exemplo emblemático na atualidade pelo uso de tecnologias aliadas as ferramentas do *microtargeting* e do *big data* na conquista dos eleitores americanos.

5.1 DEFINIÇÃO E EVOLUÇÃO DAS CAMPANHAS ELEITORAIS

É comum entender uma campanha eleitoral como o momento mais intenso da disputa entre candidatos na disputa por um cargo público. O termo campanha é derivado da palavra francesa “campo aberto” (POPKIN, 1996). Esse significado traz para a política a metáfora militar de que as campanhas são batalhas para fora dos parlamentos e sedes executivas, disputadas através de um debate público de opiniões com vistas a conquistar o voto dos eleitores.

Gomes (2004, p. 14) destaca o significado do termo campanha no sentido de tempo necessário para execução de uma determinada ação e define eleição como: “forma de procedimento, reconhecido pelas regras de uma organização pelo qual todos, ou alguns

⁶⁹ *Microtargeting* é também chamado de micro-segmentação ou micro-nicho de segmentação, consiste numa estratégia de marketing que utiliza os dados de consumo e dados demográficos para identificar os interesses dos indivíduos específicos ou grupos muito pequenos de indivíduos e influenciar os seus pensamentos ou ações. Uma meta importante de uma iniciativa *microtargeting* é conhecer o público-alvo tão bem que as mensagens são entregues através do canal de comunicação preferido do alvo.

membros da mesma, escolhem uma pessoa, ou um reduzido número de pessoas, para desempenhar um cargo com autoridade em tal organização” (GOMES, 2004, p 14).

Para Grandi *et al.* (1992), a disputa eleitoral pode ser dividida em duas grandes fases: a primeira de avaliação e posicionamento estratégico, quando ainda não há combate direto. E uma segunda, de confronto direto, essencialmente tática, na qual prevalecem os conceitos usados na guerra. Ataque, defesa, aliança são temas tão familiares aos generais quanto aos estrategistas políticos (GRANDI *et al.*, 1992, p. 117).

Alguns autores vêm às campanhas eleitorais como um dos momentos mais intensos do processo de competição política e são também o mais importante elemento para o processo eleitoral democrático (NORRIS, 2000; GOMES, 2004). Dayan e Katz (1992) denominam as campanhas eleitorais de *media events*, o que significa um acontecimento programado, limitado no tempo e no espaço, que apresenta um grupo ou uma personalidade, tem uma significação dramática ou ritual e possui ainda uma força específica que faz com que eleitores olhem para esse acontecimento.

Outros defendem que “as campanhas eleitorais devem ser lutas de ideias ou de programas, mais do que de imagens e de sensacionalismos baratos” (GOMES, 2004, p.13). Para Salgado (2012), as campanhas eleitorais fazem com que se busque *feedback* entre os organizadores e o público-alvo e são um processo que ocorre num sistema político e social marcado por diversos fatores, como o nível de desenvolvimento socioeconômico e democrático, o que inclui a liberdade de expressão e imprensa e o grau de autonomia dos jornalistas.

O que se tem em comum a esses autores é que todos consideram as campanhas eleitorais como um momento de conquista do eleitor, fazendo com que o político ou o partido use de diversas ações direcionadas e atuantes em conjunto com a comunicação e marketing para atingir suas metas de campanha na conquista da vitória nas urnas. No entanto, alertam que é preciso pensar na organização e no tempo gasto para se alcançarem os objetivos propostos. A respeito disso, Figueiredo (1986) destaca a importância da organização da campanha relacionada a datas e providências a serem tomadas durante a sua execução.

Assim, as campanhas eleitorais, podem ser vistas como esforços organizados para mobilizar e convencer por meio da informação e da persuasão e incluem diferentes elementos que se interligam, agindo uns sobre os outros, e que têm influência sobre o resultado final (SALGADO, 2012).

Em linhas gerais, o grande objetivo da campanha eleitoral é conquistar e fidelizar o eleitor, ou seja, conquistar votos e cristalizar na mente das pessoas a intenção de voto,

minimizando a possibilidade de desistência. Isso se dá pela definição de estratégias de comunicação na campanha, orquestradas por um planejamento minucioso de identificação de eleitor-alvo (target), avaliação do concorrente, pontos fortes e fracos da disputa eleitoral, segue para estabelecer estratégias e mensagem do candidato (MANHANELLI, 1988).

Assim, depois disso, é preciso definir os canais de comunicação que serão utilizados na campanha, que podem ser os diretos ou os mediados. Os canais de comunicação diretos são aqueles meios que permitem aos atores políticos comunicarem-se diretamente com o eleitorado apenas com a mediação do suporte. É o caso dos *websites* dos partidos, do *YouTube*, ou das ferramentas *on-line* que permitem a interação direta entre eleitores e candidatos – ou o seu *staff* –, como o *e-mail* ou o *Facebook*, e outros programas como o Horário Eleitoral Gratuito no Brasil, por exemplo. Os canais de comunicação mediados são aqueles que permitem aos partidos e aos candidatos transmitir a sua mensagem com a mediação dos jornalistas nos órgãos de informação de massa, isto é, a rádio, a televisão e a imprensa escrita (SALGADO, 2012).

As campanhas eleitorais são descritas por parte da literatura ora como o processo de “americanização”, ora como um processo de “modernização”. De qualquer forma, as perspectivas dos estudiosos em relação ao papel da mídia no processo eleitoral são de que as campanhas políticas sofreram alterações nos últimos tempos, acompanhando as significativas mudanças na sociedade e as inovações tecnológicas (BLUMER; KAVANAGH, 1999).

No que se referem à “americanização” das campanhas, os especialistas apontam o domínio quase absoluto do marketing político nas disputas eleitorais. Americanização é uma referência à campanha presidencial norte-americana de 1960, quando Richard Nixon e John Kennedy disputaram, pela primeira vez na televisão, frente a frente, a preferência do eleitor. O conceito de americanização é usado desde então, para referir-se descritivamente a elementos das campanhas eleitorais e a atividade profissional conectada a elas (SWANSON; MANCINI, 1996). Essa estratégia está sendo aplicada e adaptada de várias formas em outros países, inclusive no Brasil, com a importação de técnicas de marketing eleitoral que são adaptadas ao processo eleitoral brasileiro.

Para Negreiros a “americanização” das campanhas políticas tem ocorrido por duas razões:

Primeiro, devido ao grande interesse que as campanhas americanas recebem da cobertura jornalística de todo o mundo. Em segundo, pelo grande número de especialistas de campanhas que visitam os Estados Unidos para estudar e conhecer o processo eleitoral. Além disso, a publicação de livros e manuais sobre o assunto tem ajudado a espalhar os métodos das campanhas americanas para outros países; como também, essa profissionalização é apoiada pelo frequente envolvimento de

consultores políticos americanos em campanhas de outros países. Em decorrência, a disseminação destes elementos tem naturalmente sido descrita como campanha política americanizada em outros países (NEGREIROS, 2011)⁷⁰.

Nas campanhas “americanizadas”, a televisão é o veículo de comunicação por meio do qual os eleitores encontram os candidatos, ou seja, é o aparelho televisivo que passa a estabelecer uma conexão entre o representante e o representado (SWANSON; MANCINI, 1996). Por outro lado, Mancini e Swanson (1996) propõem um uso mais restrito do termo americanização - limitado ao estudo da exportação de profissionais e técnicas norte-americanos de campanha - enquanto se valem do termo modernização para descrever as transformações mais gerais das campanhas eleitorais dos outros países. Já Norris (2000) credits as transformações nas campanhas eleitorais a partir da reconfiguração da imprensa, do forte desenvolvimento da indústria televisiva e da contribuição de novos instrumentos de comunicação interpessoal, pela via da Internet e de outros meios. Para a autora, modificam-se assim as estratégias de comunicação política e eleitoral.

Nessa linha, Norris (2000) divide o desenvolvimento das campanhas eleitorais em três fases: (1) pré-moderna; (2) moderna; (3) pós-moderna. De acordo com a análise apresentada pela investigadora inglesa, os EUA e o Reino Unido são exemplos de países que, a partir de meados do século XIX até a década de 1950, viveram campanhas pré-modernas, caracterizadas por um enorme envolvimento de voluntários partidários locais que colaboravam ativamente na organização da campanha, sob coordenação dos dirigentes dessas mesmas forças políticas. O Quadro 10, elaborado por Norris (2000), destaca as principais diferenças percebidas pela autora na segmentação das três fases de campanha.

⁷⁰ Citação retirada do documento on-line. Disponível em: <http://www.achegas.net/numero/cinco/davys_negreiros.htm>. Acesso em: 20 Jan. 2014.

Quadro 10: A diferença entre as fases da campanha eleitoral

	<i>Pré-moderna</i> Metade do século XIX até anos 1950	<i>Moderna</i> Início dos anos 1960 até final dos anos 1980	<i>Pós-moderna</i> A partir de 1990
Organização de Campanha	Local e com voluntários de partidos descentralizados.	Coordenação nacional com grande profissionalização.	Coordenação nacional, mas operações descentralizadas.
Preparações	Curto prazo.	Campanha longa.	Campanha permanente.
Coordenação central	Líderes do partido.	Sedes centrais do partido, mais conselheiros especialistas.	Unidades especiais de partido para campanhas, mais consultores profissionais.
Feedback	Prospecções locais e encontros partidários.	Pesquisas de opinião ocasionais.	Pesquisas de opinião com regularidade, incremento com grupos focais e sites interativos.
Mídia	Imprensa partidária, pôsters, panfletos locais e transmissões no rádio.	Transmissões pela televisão.	Televisão fechada, emails e anúncios direcionados.
Eventos de Campanha	Reuniões públicas locais e visitas de líderes.	Gestão de notícias, conferências de imprensa diária e controle de fotos.	Extensões da gestão de notícia para rotina de política e governo.
Custos	Baixo orçamento.	Moderado.	Altos custos para consultores profissionais.
Eleitorado	Socialmente estabilizado e alinhado aos partidos.	Desalinhado social e partidariamente.	Desalinhado social e partidariamente.

Fonte: Norris (2000, p. 138) traduzido por Thauny (2014, p.22)

A fase pré-moderna se distingue bastante das fases posteriores, pois aqui ainda há formas mais diretas de comunicação entre eleitores e candidatos e uma forte tendência de identificação e adesão aos partidos.

Já na fase chamada de moderna, que corresponde à segunda metade do século XX, as campanhas eleitorais são marcadas pela televisão que substitui a campanha face a face e transforma o cenário de competição eleitoral. Este é o marco da profissionalização da campanha abordado por Mancini e Swanson (1995). Neste momento também já há uma menor identificação com os partidos. No caso brasileiro, segundo Albuquerque (1999), essa fase só ganha destaque no final da década de 1980, com o retorno da democracia. Isso está presente em 1989, nas primeiras eleições presidenciais após o regime militar. Azevedo (1998) identifica essa transformação como um processo de modernização das campanhas eleitorais no Brasil, que passam a incorporar em suas ações o uso sistemático de sondagens de opiniões, as pesquisas quantitativas e as pesquisas qualitativas, na formulação das estratégias de campanha e o uso extensivo da mídia como elemento de comunicação entre o candidato e o eleitor. Esse quadro atual das campanhas eleitorais também pode ser referenciado através do conceito de “democracia de público” (MANIN, 1995). Para Manin (1995), o surgimento da democracia de público marcaria dessa forma a passagem de um discurso mais partidariamente

ideologizado para uma democracia pasteurizada, formatada através das pesquisas de opinião e das técnicas de marketing, bem como centralizada nos meios de comunicação (na televisão).

A terceira fase de campanhas eleitorais tem seu início com o advento das NTIC's (Novas Tecnologias de Informação e Comunicação), que se acentuou no início da década de 1990. O papel central das mídias de massa em conectar os cidadãos não foi substituído, apenas complementado pelas novas tecnologias (IASULAITIS, 2012). No roll dessas novas tecnologias destaca-se a Internet, que, ao decorrer dos pleitos eleitorais, tem ganhando cada vez mais espaço e eficácia na condução das campanhas eleitorais, juntamente com as redes sociais.

A competitividade presente nos cenários eleitorais em âmbito nacional justifica os estudos de comunicação eleitoral em profundidade, principalmente no período pós-moderno de campanha (NORRIS, 2000), quando visivelmente as campanhas assimilaram processos de profissionalização, aglutinando profissionais de diferentes expertises dentro de um organograma complexo de campanha eleitoral apontado por diferentes pesquisadores da área de comunicação política (PANKE, 2013; KUNTZ, 2002; MANHANELLI, 1992; FIGUEIREDO, 1986). Os profissionais de marketing eleitoral possuem a função de moldar o candidato desde a criação de uma plataforma política até a “embalagem”, ou melhor, a conduta política, a forma de se vestir, de se expressar, tornando a sua aceitação no mercado mais palatável por meio da mídia. O candidato identifica necessidades do seu eleitorado, cria uma marca e depois procura ampliar e vender sua imagem, além de ampliar seu apoio. Assim, o marketing não muda o indivíduo, mas o molda ao mercado, da mesma forma como se molda um produto (CHAIA, 1996). Segundo Azevedo (1998), a profissionalização das campanhas eleitorais no Brasil é resultado de um processo de modernização com o uso de sondagem de opinião, uso extensivo da mídia e a personalização da campanha.

No que se refere aos meios de comunicação, é possível nessa fase também observar a fragmentação da televisão em diversos canais abertos e via satélite, o aumento das programações de rádio e a popularização da internet. Essas inovações tecnológicas impulsionaram uma série de transformações sociais e partidárias no período, o que levou a se repensarem as estruturas de campanha, principalmente no que se refere à comunicação política. Silva (2012) enfatiza a necessidade de as campanhas eleitorais pós-modernas recorrerem a plataformas interativas sem se afastarem dos canais tradicionais, destacando assim a importância da gestão dos meios de comunicação integrada para a campanha.

Além disso, há segmentação das mensagens e estudos mais aprofundados, baseando-se nas pesquisas de opinião pública, para identificar as características do eleitorado e como

atingi-lo por meio das estratégias de comunicação usadas pelos profissionais. É neste contexto que as novas tecnologias de informação e comunicação passam a ser usadas no marketing eleitoral (NORRIS, 2000). A partir dessa caracterização de Norris (2000) para as campanhas atuais é possível ampliar os usos das tecnologias da informação e comunicação em campanhas eleitorais, não somente como ambientes para divulgação de ações e mensagens de campanha.

5.2 CAMPANHAS ELEITORAIS: DA MASSIFICAÇÃO PARA A SEGMENTAÇÃO

Pela história das campanhas eleitorais, percebe-se que não é de hoje que a comunicação mantém um importante diálogo com a política. O surgimento das mídias eletrônicas – como o rádio, a televisão e o computador -, importantes meios de divulgação em massa de ideias das campanhas eleitorais, contribuem para que as informações atinjam maior número de pessoas.

Nesse sentido, reconhece-se que a mídia é um fator central da vida política contemporânea. “É ocioso alimentar a nostalgia de ‘tempos áureos’ da política, quando imperava o verdadeiro debate de ideias sem a preocupação com a imagem ou a contaminação pelas técnicas da publicidade comercial” (MIGUEL, 2002, p. 158).

Dessa forma, Miguel (2002) corrobora com os pensamentos de Veiga e Godim (2001) sobre a influência que os meios de comunicação exercem em processos eleitorais. Segundo os autores:

É bastante visível que o desenvolvimento dos meios de comunicação de massa, particularmente o rádio e a televisão, imprimiu um novo ritmo aos processos sociais e políticos. Uma de suas maiores consequências foi a de aproximar o candidato à realidade cotidiana de seu eleitorado, minimizando o papel do partido nesta mediação (VEIGA; GODIM, 2001, p.4).

Outra vantagem apontada por Veiga e Godim (2001) foi a de garantir a sincronidade das informações, ou seja, permitir que as pessoas dirijam sua atenção, ao mesmo tempo, para um único assunto, dada a força de sua penetração social. “E, se as pessoas podem ter como centro de suas discussões o mesmo assunto, ficam mais susceptíveis às pressões para a homogeneização de suas opiniões” (VEIGA; GODIM, 2001, p.4).

Entende-se, dessa forma, que os meios de comunicação, na fase pós-moderna (NORRIS, 2000), passam a ser os principais mediadores e influenciadores entre o candidato e os eleitores. Assim, em períodos eleitorais, os meios de comunicação se destacam como elo entre eleitores e seus representantes. As necessidades da população e decisões políticas são

efetivadas e legitimadas e se tornam de conhecimento público por exposição na mídia, de acordo com Lourenço (2002).

Em um estudo, Fontenelle também reconhece que a centralidade das campanhas passou a ser as telas das televisões, e sua modernização nas sociedades democráticas constitui um fenômeno mundial irreversível.

Usa-se a mídia na construção das campanhas, para persuadir e ganhar o voto do eleitor. O destaque à figura do candidato é maior do que à imagem do partido. Com a personalização das campanhas, os apelos publicitários emotivos com retóricas são mais atraentes que as propostas políticas ideológicas (FONTENELLE, 2008, p. 17).

Além disso, o HGPE constitui o principal meio de embate dos candidatos envolvidos na disputa eleitoral (VEIGA; GODIM, 2001). Existe um consenso na atualidade entre os pesquisadores da área de ciência política de que, em todas as formas de regimes políticos, especialmente os democráticos, o papel da mídia na política é de grande importância (ANTONIUTTI, 2004). Dessa forma, Antoniutti (2004) observa também que, de acordo com as pesquisas realizadas na área de comunicação política, as transformações tecnológicas e a realidade sociopolítica têm levado à redefinição de estratégias informativas e comunicativas, buscando o máximo de eficácia de seu uso. Os candidatos pensam, agem e concebem suas estratégias políticas já vislumbrando que tipo de aproveitamento essas imagens e mensagens terão por parte dos jornais, rádios, televisão (SILVA, 2002, p. 76).

Kuntz (1998) diz que um candidato pode contar com os veículos de comunicação para adequar o discurso político às suas necessidades, transmitindo e selecionando as mensagens destinadas a cada tipo de público. Reconhecendo a importância e eficácia dos meios de comunicação de massa a legislação brasileira, através da lei Nº 4.737, estipulou em 1965 o Código Eleitoral Brasileiro que institui o Horário Eleitoral Gratuito, regulamentando o acesso dos partidos à propaganda eleitoral e fiscalizando as atividades de marketing político.

Mas há que se considerar nos dias de hoje a internet como uma grande aliada nas campanhas eleitorais. A rápida expansão da rede mundial de computadores transformou esse espaço em um importante campo de disputas simbólicas e de comunicação entre a esfera política e a população, seja na administração da visibilidade (THOMPSON, 1998), na interação entre os atores políticos com a população, na mobilização política, ou ainda, na condução de novos mecanismos de realização de campanhas eleitorais, entre outras.

Assim, campanhas políticas contemporâneas têm apostado na internet como meio alternativo para veiculação eleitoral e que são próprios do ambiente virtual (GOMES *et al.*, 2009). Para Gomes (2009) existem três estágios no uso da rede pelos políticos: (i) “pré-web”,

que concentra o uso da internet no envio de mensagens eletrônicas e e-mails, sem um uso interativo dos websites; (ii) um estágio “Web”, que tem como plataforma básica de contato dos políticos com o público em geral os websites pessoais basicamente reproduzindo os conteúdos das mídias tradicionais; (iii) e uma terceira etapa iniciada com a campanha presidencial de Barack Obama, qualificada de “pós-web”, onde os websites passam a ser apenas uma ferramenta virtual entre outras empregadas pelos políticos, cumprindo a função de ser um distribuidor de conteúdo a ser reproduzido nas mídias sociais.

Como se percebe, as campanhas eleitorais passam a incorporar pouco a pouco os dispositivos da web em suas estratégias de comunicação, ampliando o escopo de ação da equipe de comunicação e marketing político, a esse tipo de marketing convencionou-se chamar de marketing digital ou marketing 2.0.

Assim a internet por sua estrutura de rede e sua prática interativa e colaborativa característica da web 2.0, foge da rotina comunicacional da mídia tradicional, caracterizada pelo fluxo unidirecional da informação, criando formas de comunicação nas quais o receptor da mensagem tem um papel mais ativo, podendo produzir e interagir com o conteúdo que vai circular na rede, de forma mais segmentada e personalizada. O termo web 2.0 é utilizado para descrever a segunda geração da internet (a primeira era chamada de web 1.0) do intenso desenvolvimento tecnológico que potencializa a criação de um cenário propício à colaboração dos eleitores. “A web 2.0 refere-se não apenas a uma combinação de técnicas informáticas, mas também a um determinado período tecnológico, a um conjunto de novas estratégias mercadológicas e a processos de comunicação mediada por computador” (PRIMO, 2007, p.2). Para Recuero (2009) diz que o maior diferencial da web 2.0 talvez seja a possibilidade da interação através de diversas ferramentas tais como sites de redes sociais, que torna o usuário um agente ativo e interativo. O que se entende é que, nessa geração da web, o usuário é a peça central porque, ao invés de receber passivamente as informações, como acontece em outros meios de comunicação, os usuários desses sites se tornam agentes produtores de informações.

Para o entusiasta da política na rede, Canavilhas (2009), é graças à internet que a política parece ter encontrado uma forma de retomar o contato direto com o eleitorado, evitando assim intermediários, como a imprensa tradicional. O autor observa que os políticos americanos desde 2004 apostam nas potencialidades comunicativas e informacionais da internet. A blogosfera foi um fenômeno em foco, com Howard Dean, candidato à nomeação democrata, que na época foi o primeiro a criar um *blog*. Esta aposta na internet permitiu-lhe abrir vias diretas no contato com o eleitorado e arrecadar 15 milhões de dólares para apoio à campanha. Da mesma forma, na corrida à Casa Branca, Bush e Kerry também apostaram no

meio, criando diários de campanha na blogosfera. Mas foi a partir da campanha de Barack Obama em 2008, que a Internet passou a ser vista como um novo elemento importante na manobra de um candidato. Obama utilizou quase todas as aplicações on-line disponíveis no momento. Além disso, criou espaços nas redes sociais, como *Obama Everywhere*, e só no Facebook chegou aos 320 mil utilizadores. Obama arrecadou ainda 28 milhões de dólares de apoio, o que lhe permitiu manter alguma distância em relação aos adversários. Desde a experiência do presidente norte-americano Barack Obama, campanhas políticas têm apostado na Internet como meio alternativo para a aplicação do marketing político e eleitoral.

No Brasil as mídias sociais passaram a ser palco das disputas eleitorais a partir de 2010. Para Sousa (2010) a expectativa pela primeira eleição pós-Obama para presidente no Brasil era enorme. Após a eleição do candidato Barack Obama nos Estados Unidos em 2008, todos os candidatos a cargos políticos passaram a dar um valor maior à internet na campanha eleitoral. No Brasil, isso não foi diferente. Vários candidatos, principalmente os candidatos à Presidência da República, usaram a internet e principalmente as redes sociais, buscando maior proximidade dos internautas e visibilidade para suas campanhas.

Algumas campanhas, como a da candidata Marina Silva (PV) para presidência da República são considerados cases de sucesso e referenciados em muitos estudos (COSTA, 2011; SOUZA, 2010). Aqui serão destacados alguns dos pontos importantes dessa campanha, considerada pioneira no Brasil no uso efetivo das mídias sociais digitais.

Assim como Obama estava em desvantagem de horário eleitoral gratuito em relação a seu principal adversário Jonh McCain em 2008, Marina Silva teve grande desvantagem em relação a seus principais oponentes à Presidência da República. Esta semelhança levou Marina Silva a buscar algumas das estratégias de sucesso de Obama na campanha de 2008 para tentar surpreender os demais candidatos (SOUSA, 2010, p.21).

Em consequência do pouco tempo de horário eleitoral no Rádio e na Televisão, a candidata do Partido Verde Marina Silva apostava na internet para se aproximar mais dos eleitores brasileiros e apresentar, detalhadamente, suas propostas de campanhas.

O desafio da equipe de internet era compensar o pouco tempo na TV com muito tempo na rede, conquistar a maioria dos eleitores via internet, mesmo que o total deles (incluindo as crianças, que não votavam, mas podiam ficar conhecendo Marina Silva e falar dela para seus pais e parentes) significasse bem menos pessoas do que o total de votantes do País (COSTA, 2011)

A campanha de Marina Silva usou todas as ferramentas da internet e tinha como objetivo transformar cada internauta que interagisse de alguma forma com a campanha num multiplicador. “Neste sentido, foi de fundamental importância a decisão de usar a rede unicamente com fins propositivos, sem atacar e sem criticar qualquer adversário político. A

‘demonização’ dos adversários não fez parte da política de Marina Silva na rede nem na mídia tradicional” (COSTA, 2011). Para comprovar o efetivo uso da rede na campanha de Marina uma pesquisa sobre a “saudabilidade” da marca Marina Silva na internet, publicada pelo semanário Meio & Mensagem, principal órgão do mercado publicitário e da indústria de comunicação, mostrou que a estratégia de alguma forma dera certo. No final de setembro, a “marca” Marina Silva carregava 84% de citações positivas na rede contra 59% de José Serra e 45% de Dilma Rousseff. Doze semanas antes, quando este acompanhamento começou, o índice de Marina Silva era de 80%, o de Serra 50% e o de Dilma 69%. Marina Silva continuou num crescendo, Serra melhorou um pouco e Dilma perdeu terreno (SOUSA, 2010).

No entanto, se as mídias sociais tornaram-se nos últimos anos um campo fértil para a divulgação de ações políticas, também passou a ser vista como um ambiente de captura de dados e informações de eleitores e que merecem mais atenção por parte dos estrategistas e especialistas em campanhas eleitorais na hora da formulação das suas ações de comunicação eleitoral.

5.2.1 O uso do marketing na política

O marketing é uma das mais antigas estratégias utilizadas no mundo, para divulgação e interesse de vendas e produtos, ou do que se pretende realizar. Mas foi somente no século XX que se tornou sistematizado, em razão da grande concorrência enfrentada pelas empresas. Nessa época os empreendedores passaram a entender que, para criar novos mercados, seria necessário desenvolver novos produtos e serviços, oferecendo maior variedade, se opondo às técnicas da época de simplesmente abastecer o mercado com o que já era produzido. Após a segunda metade do século XX com a obra “Prática de Administração de Empresas”, de autoria de Peter Drucker, o marketing passou a ser visto como filosofia, vindo a ser estudado e aplicado por empresários, executivos, consultores e acadêmicos.

No entanto, a utilização de técnicas de marketing de produtos na arena política é um fenômeno contemporâneo. A migração de conceitos e práticas do universo empresarial para a política tem-se tornado, no mundo todo, a tendência predominante nas últimas décadas. Kotler (1975) considera que o marketing político é comparável ao marketing comercial seja através de seus conceitos e ferramentas, seja através dos modelos de comportamento do consumidor. A busca pelo poder exige dos candidatos o mesmo tipo de planejamento, pesquisa e posicionamento que existe na área comercial. Para Figueiredo (1995), marketing político e marketing de produtos são conjuntos de ações (de partidos e empresas) que querem atrair os “consumidores” (eleitores).

É evidente, no entanto, que entre o marketing político e o de produtos existem diferenças. Segundo Tomazelli (1988) “um candidato é um produto que pensa, fala e age por conta própria, diferentemente de um produto comercial específico que não pode modificar-se e adaptar-se a novas situações”. O autor diz ainda que um homem não pode ser planejado como um produto, pois ele tem antecedentes, personalidade e é interpretado por suas ações e palavras.

O marketing político engloba ferramentas de publicidade, jornalísticas, técnicas teatrais, musicais, sensoriais, humorísticas e de psicologia. Além disso, são usadas pesquisas, análises estatísticas e qualitativas ao máximo. Tudo é planejado e unido a um profundo entendimento do ser humano e seus anseios.

Para Torquato (2014), o marketing político tem uma pré-história. O autor lembra que esse conceito nasceu no ano 64 a. C., quando Quinto Túlio Cícero enviou ao irmão o grande tribuno e advogado Cícero – protagonista de episódios marcantes por ocasião do fim do sistema republicano, e da implantação do Império Romano uma carta que é considerado o primeiro manual organizado de marketing eleitoral da história.

A respeito disso, Torquato ressalta:

Ali, Quinto Túlio orientava Cícero sobre comportamentos, atitudes, ações e programa de governo para o consulado, que era o pleito disputado, sem esquecer as abordagens psicológicas do discurso, como a lembrança sobre a esperança, esse valor tão marketizado no Brasil que tem se firmado como base para eleições vitoriosas: “Três são as coisas que levam os homens a se sentir cativados e dispostos a dar o apoio eleitoral: um favor, uma esperança ou a simpatia espontânea”. E passava a discorrer sobre cada valor (TORQUATO, 2014, p.21).

O autor ainda observa que esses escritos demonstram como desde a época de Cícero as redes sociais já eram discutidas, quando Túlio recomendava a “atenção para a cidade inteira, todas as associações, todos os distritos e bairros”. Segundo o autor, o que mudou dessa época para os dias atuais foram os meios de divulgar e expandir um discurso político, cujos vértices sempre apontaram para questões de princípios, valores, qualidades, posições entre o bem e o mal, enaltecimento de perfis, desigualdade social.

Já no mundo contemporâneo a primeira vez que um candidato utilizou uma grande empresa de comunicação numa campanha presidencial foi em 1952, nos EUA, com o general Dwight Eisenhower (1890-1969). O general contratou a agência publicitária BBDO para auxiliá-lo na campanha presidencial americana daquele ano. A principal função da empresa de comunicação era adaptar a linguagem de Eisenhower à mídia eletrônica disponível à época – rádio e televisão (FIGUEIREDO, 2000).

No Brasil, Prudente de Moraes na década de 1894 foi quem inaugurou o conceito de comunicação e marketing político no seu governo como presidente do Brasil. Uma das suas ideias iniciais foi usar jornais para alcançar a população e mostrar aos eleitores seu trabalho já desenvolvido como “presidente de São Paulo” - governador da época. Mas, foi na campanha eleitoral de 1989, do candidato Fernando Collor de Mello, que as ferramentas de marketing tiveram maior destaque, inaugurando um novo tempo para a comunicação e propaganda eleitoral.

Nos dias atuais, o papel da mídia na política torna-se importante a cada dia, pois a sociedade experimenta um novo mundo de informações e inovações que propiciam o uso das ferramentas de comunicação. E nesse cenário o político usa todos os artifícios e “artimanhas” para convencer os seus eleitores disso e conta com profissionais de comunicação e marketing para divulgar suas ações e intenções da melhor forma possível e pelos canais de comunicação mais apropriados. Para isso, o marketing político conta com diversos recursos de monitoramento da opinião pública tais como pesquisas *survey* e qualitativas, que informam os políticos sobre as expectativas e visões dos eleitores (RIBEIRO, 2002; KOTLER, 1999). Ele possibilita, aos candidatos, adaptar, rever, modificar ou reforçar suas estratégias de apresentação, seus discursos e suas práticas, na cena pública.

Para Figueiredo:

O marketing político é um conjunto de técnicas e procedimentos cujo objetivo é avaliar, através de pesquisas qualitativas e quantitativas, os humores do eleitorado, para, a partir daí, encontrar o melhor caminho para que o candidato atinja a maior votação possível (FIGUEIREDO, 2000, p.14).

A partir do conjunto de informações disponíveis, o marketing manipula os dados, e os políticos se esforçam, através do marketing político, para estabelecer um elo simbólico com o eleitorado, recorrendo a saberes, normas, valores e universos de referência supostamente partilhados.

5.2.2 Do marketing eleitoral para o marketing individual

As campanhas políticas atuais têm experimentado diversos tipos de marketing em suas estratégias. Na prática, o que ocorre geralmente é que as ações de marketing restringem-se ao processo exclusivamente eleitoral e perduram apenas até o momento do voto do eleitor.

O marketing eleitoral trata de uma estratégia voltada para o eleitor, com o objetivo de fazer o partido ou candidato vencer uma determinada eleição. É somente um instrumento de comunicação em campanhas. Abrange todas as técnicas de comunicação disponíveis no mercado. É uma estratégia mais global, que inclui – além da comunicação – a pesquisa, o

planejamento estratégico, o uso de especialistas e a articulação de um conjunto de decisões e ações necessárias para a atuação política de modo geral, além da eleitoral estrito senso (RICHERS, 2007). E nesse contexto, entende-se que as estratégias das análises de dados (*big data*) são integradas a esse conjunto de ações que se tornam importantes para a disputa eleitoral.

Dentro das disputas eleitorais, o fato crucial no uso de estratégias de marketing é o tempo. Assim, os candidatos precisam desenvolver um trabalho contínuo e sistemático de formação de imagem, objetivando aumentar seu poder de influência sobre o eleitor na decisão de voto. E nesse sentido o uso da tecnologia e ferramentas podem ajudar os profissionais tanto na coleta de informações quanto na disseminação de mensagens do candidato.

Por fim, é importante ressaltar que o marketing é um conjunto de planos e ações desenvolvidos para influenciar os eleitores no período eleitoral, mas não é ele quem define uma eleição. Considerar que o marketing eleitoral pode eleger qualquer político somente através de suas ações de comunicação é menosprezar o papel do eleitorado que tem se apresentado mais maduro (SILVA, 2002) e também cada vez mais racional no processo de escolha política. Nesse sentido, as campanhas têm se mostrado cada vez mais científicas e mais baseadas em resultados de análises de dados, como afirmam Moura e Pereira⁷¹ (2014). Daí a importância de dar “ouvidos” para que os eleitores e cidadãos esperam e querem de seus governantes e representates.

Em tempos de internet e mídias sociais, o que prevalece principalmente em campanhas políticas brasileiras desde o case da campanha eleitoral americana de Barack Obama em 2008, é o chamado marketing político 2.0, ou marketing eleitoral 2.0. Para alguns autores a internet surge para o marketing político explorando espaços de visibilidade acessíveis a qualquer hora e a qualquer lugar.

Diferentemente dos meios de comunicação de massa, a internet oferece grande variedade de informações, e não apenas material de origem oficial. Segundo Gomes (2009), reduz custos de participação e permite envolver diferentes parceiros de interlocução. Para o marketing político, a rede proporciona um meio diferente do Horário Eleitoral Gratuito, santinhos, panfletos, cartazes e demais propagandas impressas. A internet inaugura um espaço no qual o “público e os políticos podem comunicar-se, trocar informações, consultar e debater de maneira direta, rápida e sem obstáculos burocráticos” (GOMES, 2009, p. 277).

⁷¹ Maurício Moura e Renato Pereira fizeram essas afirmativas durante as entrevistas para a autora desta tese.

Entende-se que hoje uma campanha eleitoral não pode mais ser tornar um jogo de acertos e erros. Antes de partir para uma disputa é necessário um estudo aprofundado do momento eleitoral, de uma estruturação correta, de um desenvolvimento competente e de um acompanhamento atento, capaz de antever facilidades ou problemas de percurso, para acelerar ou frear nos momentos adequados as ações de uma campanha política. Nesse contexto, o foco no comportamento do eleitor é indispensável. Explorar informações sobre os eleitores é imprescindível para a realização das estratégias de marketing eleitoral. Como bem afirma Manhanelli (1988), nas estratégias aplicadas em campanha eleitorais, o mais importante são as informações, e “a transformação das informações adquiridas em ação efetiva é a famosa estratégia. Assim, quanto mais precisas forem as informações usadas para se montar uma estratégia, menor chance de erro na estratégia de ação proposta” (MANHANELLI, 1988, p.16).

Outro ponto importante é de que o marketing eleitoral cada vez mais se utiliza de outros tipos de marketing para fidelização da candidatura junto aos seus eleitores, porque nenhuma disputa eleitoral para o candidato termina no dia da votação, ele sempre tem a intenção da continuidade política e de manter um relacionamento mais duradouro com seus eleitores. Com esse propósito, o marketing eleitoral vai buscar estratégias no marketing de relacionamento e no marketing *one to one*, muito usado pelas empresas em suas áreas de negócios. Assim, segundo McKeena (1991) define, marketing de relacionamento⁷² ou pós-marketing significa construir e sustentar a infraestrutura dos relacionamentos de clientes.

Para Peppers e Rogers (2001, p.10), “o processo de implementação de um programa de marketing *one to one* pode ser pensado como uma série de quatro passos básicos: identifique seus clientes; diferencie seus clientes; interaja com seus clientes; personalize”. Adaptado às estratégias eleitorais esse marketing tem como estratégias: (1) **Identificar**, o que consiste em enxergar seus eleitores, colhendo dados e principalmente mensurando suas ações nos diversos pontos de contato com eles, seja on-line ou off-line; (2) **Diferenciar**, o que consiste em ter uma ferramenta adequada que lhe permita segmentar eleitores; (3) **Aproximar-se**, ou seja, iniciar um diálogo contínuo com estes eleitores de maneira a

⁷² O Marketing de relacionamento diz respeito a uma nova postura das empresas para com seus clientes, ele confirma uma mudança de foco que saiu da venda dos produtos para a satisfação das necessidades dos consumidores. São esforços de objetivos a longo prazo. Com esse redirecionamento de foco que saiu da venda para o relacionamento, as compras imediatas já não são interessantes para as empresas, o que se busca são contatos mais prolongados que sejam capazes de unir ainda mais clientes e empresas independentes do volume de compras realizado pelos primeiros. Mas para isso as empresas investem bastante em mecanismos capazes de detectar os desejos e necessidades de seus consumidores para assim, oferecê- los exatamente aquilo que os seus consumidores buscam. É nesse ponto que o elo de valor se forma e a antiga relação comercial se transforma em uma relação de amizade (MCKEENA, 1991).

incrementar constantemente o conhecimento sobre cada um deles, especialmente quanto às suas necessidades, desejos e comportamento; (4) **Personalizar**, adaptando suas ideias e projetos de governo de acordo com a individualidade de seus eleitores.

Assim como no meio comercial, o marketing individualizado e personalizado permite um grande diferencial competitivo, principalmente porque desta forma a campanha política pode entender melhor o seu ambiente eleitoral e antecipar-se às ações dos candidatos adversários. Nesse contexto, as contínuas mudanças ambientais influenciam a maneira de exercitar a política, principalmente em tempos de mídias participativas, como as redes sociais digitais. Assim, entende-se nesta tese que a tarefa do marketing no segmento da política torna-se extremamente complexa, arriscada e cada vez mais necessária para reduzir os riscos do processo político. E nesse cenário, a informação sempre se mostra como uma importante condutora de todos os processos numa campanha eleitoral.

5.3 A INFORMAÇÃO E A PESQUISA NO AMBIENTE ELEITORAL

Desde os primórdios das relações de troca, a informação já se mostrava uma peça vital. Comerciantes dos tempos mais remotos utilizavam as “dicas” que dispunham para precificar seus produtos e propor escambos – roupas, modo de falar e se portar, habilidade nas interações do interlocutor etc., tudo isso eram informações essenciais que ajudavam o comerciante a montar suas estratégias rudimentares de persuasão e vendas.

Para ilustrar essa discussão utiliza-se também o exemplo das metodologias seminais de estratégia, de Sun Tzu, encontradas na obra considerada o mais antigo relato escrito sobre estratégia no mundo: *A Arte da Guerra*⁷³. O que hoje é conhecido por estratégias aplicadas principalmente pelo marketing, nasceu muitos anos antes de Cristo, nos escritos do autor oriental (TZU, 1983). “Para vencer, deve conhecer perfeitamente a terra (a geografia, o terreno) e os homens (tanto a si mesmo quanto o inimigo). O resto é uma questão de cálculo. Eis a arte da guerra”. (TZU, 1983, p. 6). Em outra parte do livro que vale a pena destacar e que se relaciona com a importância da informação e do conhecimento são os cinco fatores principais, que devem ser o objeto de nossa contínua meditação e de todo o nosso cuidado, como fazem os grandes artistas ao iniciarem uma obra-prima. Eles têm sempre em mente o

⁷³ O livro foi traduzido pela primeira vez pelo missionário jesuíta Amiot em 1772, sendo a primeira versão que se conhece no Ocidente. Embora as táticas bélicas tenham mudado desde a época de Sun Tzu, esse tratado teria influenciado, segundo a Enciclopédia Britânica, certos estrategistas modernos como Mao Tsé-tung, em sua luta contra os japoneses e os chineses nacionalistas.

objetivo a que visam, e aproveitam tudo o que veem e ouvem, esforçando-se para adquirir novos conhecimentos e todos os subsídios que possam conduzi-los ao êxito. “Se quisermos que a glória e o sucesso acompanhem nossas armas, jamais devemos perder de vista os seguintes fatores: a doutrina, o tempo, o espaço, o comando, a disciplina” (TZU, 1983, p.13).

Nesse sentido, as estratégias de marketing são majoritariamente fundamentadas em inteligência, ou seja, informações sobre o inimigo, o local de batalha, as condições do tempo, entre outros fatores do cenário no qual se travará a disputa. Talvez o principal ativo de uma estratégia bem sucedida seja a capacidade com que o estrategista consegue, sistematicamente, coletar, analisar e utilizar informações para auxiliar suas tomadas de decisões, que devem sempre ser rápidas, certas, com custos baixos e com alta probabilidade de sucesso. No mundo dos negócios é assim e na política não é diferente, principalmente em épocas eleitorais.

Segundo Lima, o ambiente eleitoral é cheio de informações relevantes para uma campanha. No marketing eleitoral cabe ao Sistema de Informação em Marketing Eleitoral (SIME) fornecer as informações mais importantes que vão conduzir a campanha política.

Um SIME pode ser definido como um complexo estruturado e interagente de pessoas, máquinas (objetos) e procedimentos que se destinam a gerar um fluxo ordenado de informações relevantes, coletadas interna e externamente à campanha, para uso na tomada de decisões em áreas de responsabilidade do candidato e de seus assessores mais próximos ou conforme delegação de poderes distribuídos entre o staff da campanha e seus auxiliares voluntários (LIMA, 1988, p. 34).

Dessa forma, o SIME é responsável por transformar dados brutos sobre uma determinada situação ou problema em informações úteis para uma melhor compreensão e solução para a campanha eleitoral. E isso se reflete de forma decisiva na adequação mercadológica, estratégica e operacional da campanha em relação ao meio ambiente eleitoral.

Lima (1988) destaca que a coleta de dados junto às fontes pode ser feita de duas maneiras: através de uma busca, visando atender a uma solicitação de informação específica, ou através da monitoração do ambiente, obtendo dados esparsos, mas potencialmente interessantes para a campanha, contidos em jornais, revistas, programas de tevê, rádio, catálogos de produção de teses universitárias, boletins de sindicatos, observação direta etc. A centralização deste tipo de coleta no SIME economiza tempo do candidato e de assessores importantes em outras áreas. Por outro lado, nada adianta ter os dados se eles não estão em ordem. É necessário sistematizá-los através de regras de indexação que permitam sua rápida localização, bem como armazená-los da maneira mais adequada possível. Na fase de análise dos dados e das fontes são determinados o seu grau de confiabilidade e precisão e quais

informações o cruzamento dos dados permitem obter. Com a avaliação das informações disponíveis é possível identificar as ameaças e oportunidades, bem como sugerir procedimentos a serem adotados frente às mesmas.

Lima (1988) diz ainda que SIME deve ser capaz de disseminar adequadamente as informações obtidas/construídas, procurando emitir relatórios precisos, contendo a quantidade necessária de informação, de acordo com as solicitações feitas por candidatos e assessores. Cabe nessa discussão ilustrar este trabalho com um case de uso de informações para direcionar as estratégias de campanha eleitoral.

5.3.1 Case Basílio Villani

Como exemplo prático do quanto é importante o uso de informações em uma campanha eleitoral, destaca-se aqui uma campanha para deputado federal⁷⁴. Importante ressaltar que a busca de informações nas regiões geográficas em que o político tinha sua representatividade era feita de forma organizada, atualizada e segmentada. A campanha em destaque teve como assessora de comunicação a autora deste trabalho.

De acordo com o ex-assessor parlamentar, Geraldo Siqueira (2015), a base de dados que sustentava as diretrizes da campanha vinha de diversas fontes tanto particulares como públicas: partido político; prefeituras; órgãos públicos municipais; associações de moradores dos diversos municípios de atuação parlamentar; pesquisas encomendadas de institutos; IBGE; Tribunal Superior Eleitoral (TSE); Tribunal Regional Eleitoral (TER); banco de dados do Gabinete Brasília e Curitiba do parlamentar, onde eram feitos registros por meio de cadastros manuais, toda vez que uma liderança política fazia visitas.

“Parte dos dados e informações eleitorais eram coletadas pelos assessores políticos nos trabalhos de campo, que eram realizadas em cada visita feita no município de representação do parlamentar. Dessa forma, tínhamos como saber quem eram os principais atores políticos mais importantes e relevantes para a campanha na hora da busca de apoio político em cada região do Estado” (SIQUEIRA, 2015).

Na época existia no gabinete um investimento grande na parte de informatização e digitalização dos dados e informações políticas. Uma equipe era responsável pelo setor de coleta, digitação e armazenamento de dados e cuidava de toda a parte de informatização do gabinete. Os dados coletados eram armazenados no Sistema Gerenciador de Banco de Dados

⁷⁴ O candidato aqui que é usado como exemplo é o ex-deputado federal Basílio Villani (PSDB) do Estado do Paraná, que hoje é aposentado como bancário e deputado. Basílio Villani é formado em Economia pela FAE (Faculdade de Administração e Economia) de Curitiba e fez carreira como bancário no antigo Grupo Bamerindus, onde foi escolhido pela empresa para representar a classe na Câmara dos Deputados em 1986. Foi um dos parlamentares que participou da reformulação da Constituinte Federal de 1988. No dia 29 de agosto de 1992 Villani foi um dos 33 parlamentares que votaram contra o *Impeachment* do ex-presidente Fernando Collor de Mello (PRN).

(SGBD) dBase II. O sistema também gerava consultas e relatórios. A mineração de dados não existia no sistema gerenciador, era toda feita de forma manual. Os bancos de dados eram abastecidos com diversas informações eleitorais como número de partidos, partidos políticos, seções eleitorais no estado do Paraná, número de eleitores por região e município.

O parlamentar era considerado pelos jornalistas e especialistas em política um político pragmático com objetivos muito bem definidos. O principal deles era garantir a sua eleição sempre baseada em dados reais, matematicamente calculados pelo custo-benefício do que gastaria na campanha em cada localidade (município) em relação aos votos que receberia e quando a campanha não correspondia aos benefícios, as estratégias políticas eram reformuladas, a fim de não haver desperdícios de verbas naquele lugar sem resultados expressivos. O ex-deputado lembra de como eram tomadas as decisões de campanha a partir das informações e dados que obtinha em suas bases de atuação.

“A cada eleição era feita uma avaliação do custo de cada voto que eu recebia e isso servia sempre para as futuras comparações e estratégias de ações políticas. Na próxima eleição sempre buscava os votos mais baratos. Olhava uma região e dizia: aqui o voto é muito caro. Eu trabalhava por metas. Estimulava o número de votos que queria e via o quanto precisava em cada município. Não pretendia ser o mais votado. Queria apenas o que todo político queria no cargo de deputado, garantir a minha reeleição” (VILLANI, 2015).

E quando as metas não eram atingidas em determinado município, o parlamentar mudava os planos e ações. *“Pode ter morrido um cabo eleitoral forte. Zebrou, então tinha que transferir o meu cirquinho. Pegava minha lona e levava para montar em outro lugar”* (VILLANI, 2015). Todas as ações da campanha eram direcionadas, a partir da meta voto⁷⁵ por região geográfica como demonstra o Quadro 11.

Quadro 11: Estruturação da campanha com dados eleitorais

		ORGANIZAÇÃO POLÍTICA VILLANI					Situação			Objetivo
Município		Reg.	Eleitores	1.986	1.990	1.994	P.M.	P.S.	M.V.	1.998
361	São José dos Pinhais	Ctba	96.493	1.270	2.283	174				BS
362	São Manoel		2.094							
363	São Mateus do Sul	Rio	23.262	108	2	26				
364	São Miguel do Iguaçu		17.330	45		4				
365	São Pedro do Iguaçu		5.929							
366	São Pedro do Ivaí		7.742	4						
367	São Pedro do Paraná		2.509	4						
368	São Sebastião da Amoreira		6.848	1		6				
369	São Tomé		3.949	4						
360	Sapopema	Jac.	5.672	38	44	793				RSSS
361	Sarandi		39.599	11	2	12				
362	Saudade do Iguaçu	Jac.	3.292			10				
363	Senzêles	Jac.	10.873	141	1.799	1.881				RBSS

Fonte: Campanha política do Deputado Basílio Villani (1998)

⁷⁵ Meta voto era o termo usado na campanha para designar o número de votos necessários que o candidato deveria alcançar em determinada localidade ou região da sua atuação para ser eleito (SIQUEIRA, 2015).

Este quadro representa a distribuição da campanha por municípios e regiões do Estado do Paraná. Mostra o número de eleitores atualizados na época (1998), como o candidato obteve seu desempenho em cada eleição passada (1986, 1990 e 1994) e ainda como eram as negociações políticas para a eleição atual, demonstrada por meio da situação da campanha e por fim apresenta o número (por meio de códigos) de votos necessários e esperados naquela localidade. Os cálculos para a meta voto do parlamentar em cada cidade e região eram feitos com bases matemáticas e estatísticas a partir do Quociente Eleitoral⁷⁶.

Siqueira (2015) lembra ainda que essas informações e dados eleitorais eram levantados e coletados constantemente, e a base de dados era atualizada com variáveis como nome, sexo, idade, endereço, telefones, profissão, cargo político, partido político, números de votos feitos nas últimas eleições, coligações, entre outros.

O banco de dados era dividido em três grandes listas chamadas de grupo A, grupo B, grupo C. O grupo A era constituída de informações e dados referentes a prefeitos, vice-prefeitos, vereadores, legenda e partidos de apoio que haviam vencido as eleições. No grupo B ficam armazenados os dados sobre candidatos e lideranças de outras legendas e perdedores das eleições e o grupo C ficam armazenados informações eleitorais, como número de eleitores, zonas eleitorais e a votação em cada eleição por região e municípios.

As análises de dados eram feitas ainda de forma muito manual, praticamente a “olho nú”. As informações que vinham das análises dos dados realizadas pelos profissionais da área de Tecnologia da Informação (TI), passavam sempre por uma análise mais apurada pela equipe política e de comunicação. As informações retiradas dessas análises serviam para diversas finalidades, entre elas direcionar todas as ações do candidato em momentos das campanhas eleitorais, bem como para direcionar as ações de comunicação e marketing⁷⁷ da campanha.

A partir dessas análises eram estabelecidos os direcionamentos das mensagens que eram veiculadas tanto em meios de comunicação tradicional como em canais de comunicação direto (telefonemas, porta a porta, mala direta), que eram desenvolvidos para cada cidade de atuação parlamentar. Antes de qualquer ação junto aos eleitores ou mesmo lideranças, eram realizadas as análises de segmentos de atuação do deputado.

⁷⁶ Quociente eleitoral é, em conjunto com o quociente partidário e a distribuição das sobras, o método pelo qual se distribuem as cadeiras nas eleições proporcionais brasileiras (cargos de deputado federal, deputado estadual ou distrital e vereador). De acordo com o artigo 11 do Código Eleitoral (Lei nº 4.737/65), se nenhum partido ou coligação alcançar o quociente eleitoral, até serem preenchidos todos os lugares, serão considerados eleitos os candidatos mais votados.

⁷⁷ A autora desta tese era responsável por todas as ações de comunicação do político durante o mandato e em épocas de disputas eleitorais.

Como abordam os especialistas em campanhas eleitorais, ao iniciar uma campanha eleitoral a equipe de comunicação tem que ter em mente que as primeiras ações a serem efetuadas devem ser no sentido de consolidar o segmento que apóia o candidato, direcionando todo trabalho inicial a este segmento e suas características.

Segundo Gaudêncio Torquato (2014) em uma campanha eleitoral definir segmentos é tarefa importante. A localização se completa com a identificação dos comportamentos do eleitor-alvo. Pois, o conhecimento do eleitor ensejará tópicos para ajustes de programas e mensagens no decorrer da campanha. Da mesma forma Manhanelli observa que:

Em campanhas proporcionais, torna-se vital a detecção do segmento que apóia o candidato, pois é candidato quem tem atrás de si um segmento que o considera líder e o apóia para que tenha poder político e lute pelas ideias e reivindicações deste segmento” (MANHANELLI, 1988, p. 33).

Dentro dessa linha de raciocínio, as estratégias iniciais da campanha do parlamentar, analisava desde o comportamento desse liderado até suas necessidades mais básicas. Assim, seguindo o que diz Manhanelli (1988), a campanha fazia reuniões com as lideranças locais, o candidato e os assessores com cada segmento da sociedade divididos em: Movimento Jovem, Movimento Feminino e Movimento Terceira Idade. Nesses movimentos eram realizados debates para ouvir as necessidades dos moradores locais e as expectativas de cada segmento com relação ao candidato e aí eram apresentadas as propostas do candidato para aquela localidade⁷⁸.

Com esse exemplo prático fica evidente que mesmo usando um processo manual e lento, a coleta de dados e de informações para a campanha eleitoral eram fundamentais para dar o direcionamento da campanha e para a tomadas de decisões tanto da equipe quanto do próprio candidato. E que a busca contínua por informações e dados eleitorais não é uma preocupação apenas de grandes campanhas eleitorais da atualidade. O que muda na verdade é a forma de trabalho e as ferramentas para a coleta dessas informações, que hoje se tornaram mais práticas e rápidas com o surgimento da internet e de tecnologias, que são cada vez mais incorporadas aos processos eleitorais.

5.3.2 O papel da pesquisa na campanha eleitoral

Para Ferraz (2010), dentre todas as transformações ocorridas nas campanhas eleitorais, considera a mais importante à inclusão da pesquisa política, no seu núcleo central de

⁷⁸ A internet nessa época estava na sua fase embrionária e pouca coisa era feita de modo digital, assim a coleta de dados era feita de forma analógica e diretamente nas fontes.

estratégias e decisão. Nenhuma ação de comunicação é feita antes de rever os índices de aceitação e rejeição dos candidatos. Da mesma forma Chalegre *et al* (2006, p. 4) afirma que a complexidade do ambiente que hoje se defrontam os partidos e candidatos exigem que eles dediquem mais tempo e esforços para um maior conhecimento dos componentes eleitorais, caso desejem manter-se competitivos no mercado político.

Assim, para influenciar os eleitores, saber quais conquistar e se posicionar melhor é necessário ter um conhecimento prévio do comportamento dos eleitores: quem são eles, o que pensam e como tomam suas decisões. “A decisão dos segmentos a serem conquistados pode ser tomada com mais segurança se estiver respaldada por pesquisas sérias e abrangentes do comportamento dos eleitores” (CHALEGRE *et al.*, p.3, 2006). Da mesma forma pensa Figueiredo quando diz que:

Qualquer estratégia de marketing político necessita estar amparada em pesquisas qualitativas e quantitativas, a partir das quais se definem a postura do candidato, os temas da campanha, as propostas que serão apresentadas aos eleitores, a forma de apresentá-las, quais segmentos do eleitorado devem ser priorizados e de que maneira, quais apoios são importantes, se é conveniente atacar ou não os outros candidatos etc (FIGUEIREDO, 2000, p. 20).

As pesquisas qualitativas também podem ser chamadas de motivacionais, que segundo Figueiredo (2000) são importantes na definição das linhas de campanha dos candidatos. Entretanto, a técnica mais utilizada em campanhas eleitorais são as dos grupos de discussão *focus groups* (grupos focais). “Nesse tipo de levantamento, muito utilizado nas pesquisas de mercado, recrutam-se grupos de eleitores (normalmente de 8 a 10) que tenham características sócio-econômicas semelhantes e referenciais de ordem políticas relevantes ao levantamento” (FIGUEIREDO, 2000, p.21). A pesquisa leva em consideração também os critérios que ordenam os eleitores nas classes A, B, C, D e E, levando-se em conta aspectos educacionais e capacidade de consumo. As informações levantadas nesse tipo de pesquisa são utilizadas para dar o direcionamento inclusive do discurso e dos programas eleitorais.

No entanto, alguns autores (KUNTZ, 1982; MANHANELLI, 1988; CHALEGRE *et al.*, 2006) apontam as limitações e dificuldades das pesquisas de marketing em campanhas eleitorais. A primeira delas é a restrição temporal. Para Chalegre *et al.* (2006) o tempo que existe entre o lançamento de uma campanha e a data de uma eleição é bastante curto, restringindo sensivelmente o raio de ação das pesquisas.

Mesmo se fosse decidido realizar pesquisas bem antes do início das campanhas eleitorais, ainda assim haveria restrições devido à grande dinâmica do ambiente político, onde um determinado acontecimento pode mudar completamente a decisão de voto (CHALEGRE *et al.*, 2006, p.5).

Segundo esses autores essa restrição pode ser confirmada pelo depoimento de outro autor (KUNTZ, 1982, p. 92), que diz que o grande ponto fraco de uma pesquisa é a sua efemeridade, pois ela retrata situações ou fornece informações existentes no exato momento em que é realizada. Numa disputa eleitoral, o ambiente é profundamente dinâmico e passível de radicais transformações em curtíssimo espaço de tempo, enquanto a pesquisa é sempre estática.

Manhanelli (1988) também contribui para essa linha de pensamento dizendo que a falta de tradição do uso mais amplo e profundo das pesquisas, é a postura dos políticos quanto usuário das mesmas, ou seja, a dificuldade que os políticos têm de definir de forma clara e precisa os objetivos das pesquisas que eles contratam, acaba sendo um limitante para o aprimoramento da prática da pesquisa eleitoral no Brasil. Segundo esse autor, sem o conhecimento claro do que o contratante deseja, uma pesquisa, mesmo com o rigor metodológico, pode ser de pouca utilidade para o candidato.

Tomazelli também aponta outro problema a respeito disso.

O problema para executar um bom plano de pesquisa é o seu relativo alto custo, que deverá ser incluído no orçamento da campanha, e, por isso, muitas vezes acaba sendo relegada a segundo plano. Os candidatos preferem gastar este dinheiro em carros, panfletos e cabos eleitorais (TOMAZELLI, 1986, p.30).

Para o autor isso pode ser um desperdício de dinheiro. Segundo Tomazelli (1986, p. 30), os candidatos não percebem que podem estar jogando dinheiro fora, se as suas estratégias não estiverem corretas, se não tiverem embasamento nas reais necessidades e aspirações da população.

Chalegre *et al.* diz ainda que:

O tratamento dispensado pelos meios de comunicação às pesquisas eleitorais pode ser considerado outro fator que dificulta o amadurecimento das pesquisas no Brasil, contribuindo para sua desvalorização perante os políticos, os quais, normalmente, não possuem conhecimento das técnicas utilizadas. Nesta declaração, fica claro que normalmente, as pessoas que não são especialistas no assunto não entendem e não sabem analisar o resultado de uma pesquisa, eles observam apenas os dados brutos. Essas são as principais variáveis que podem dificultar ou limitar a evolução das pesquisas eleitorais no Brasil. Resta agora, que pesquisadores e estudiosos sobre esse assunto, possam se interessar e dedicarem algum tempo para a prática das pesquisas no cenário político, difundindo cada vez mais os benefícios que elas podem oferecer (CHALEGRE *et al.*, 2006, p.6).

Acrescentam-se a isso também críticas feitas por especialistas aos institutos de pesquisas tradicionais nos últimos anos eleitorais sobre os diferentes resultados das pesquisas de intenções de votos em relação aos resultados finais das urnas. Esse foi o caso vivenciado nas últimas eleições de 2014 no Brasil. Na eleição presidencial e também para o governo de

alguns estados como Rio de Janeiro, Bahia e Rio Grande do Sul, o eleitor foi surpreendido com resultados bem diferentes daqueles divulgados pelos institutos de pesquisas às vésperas do pleito de 5 de outubro (primeiro turno).

Nesse contexto, vale ressaltar que, de forma geral, mesmos os candidatos mais experientes, não analisam de forma mais abrangente as pesquisas de opiniões. Eles geralmente, só querem saber quem está na frente, quem subiu e quem caiu. Os dados de uma pesquisa quantitativa devem ser analisados juntamente com o quadro político atual e as pesquisas qualitativas irão ajudar a explicar a motivação. O que na verdade com o uso cada vez maior das redes sociais digitais pelos cidadãos, é importante estar atento a tudo o que acontece nesses ambientes.

Nesse sentido, o *big data* se mostra na avaliação de especialistas (OLIVEIRA, 2013) como uma opção agregado as demais formas de pesquisas de marketing para aprimorar a forma de se fazerem pesquisas eleitorais no Brasil.

5.3.3 As mídias sociais como estratégias de pesquisa eleitoral

Antes do advento das mídias sociais digitais, as pesquisas sobre os processos comunicacionais utilizavam métodos empíricos baseados em amostras obtidas a partir de questionários, entrevistas e vários outros processos que permitiam compreender as características de um determinado universo. No entanto, nos dias de hoje, qualquer pessoa conectada pela internet pode ser um formador de opinião em potencial e tem condições de atingir em poucos segundos uma audiência muito maior que os meios de comunicação tradicional (Rádio, TV, Jornal).

Segundo Real (2013), uma simples amostra pode não ser suficiente para analisar o fluxo de informação. O volume de dados tornou-se muito maior, mais dinâmico e as pessoas passaram a se expressar espontaneamente e com muito mais intensidade sem terem de ser questionadas para dizer o que pensam sobre um determinado assunto (REAL, 2013, p.2).

As campanhas políticas, aos poucos, aderiram à tecnologia da informação e a toda inovação permitida por ela. Para Torquato⁷⁹ as mídias sociais mudam a cultura de participação dos brasileiros no processo político. Segundo ele:

Agora existe a opinião pública virtual, que é muito influenciada pelo que circula na internet. Nunca se viu tanta propagação de mensagens de interesse político na internet: se acontece um escândalo, uma votação polêmica em Brasília,

⁷⁹ Entrevista concedida a Revista Veja. Disponível em: <http://veja.abril.com.br/noticia/brasil/eleicao-2010-internet-redes-sociais>. Acesso em: 14 Jan. 2015.

imediatamente as pessoas começam a se manifestar nos blogs e *Twitters*. (TORQUATO, 2009)

Para o autor não é mais possível que os políticos desconsiderem o que os cidadãos estão falando nas mídias sociais. A internet hoje é um meio de comunicação relevante para a formação da opinião pública e para a expressão de opiniões, realização de discussões e até mesmo de mobilizações políticas. Compactua da mesma opinião Brandão ao dizer que “a Internet oferece a possibilidade de se tornar um espaço midiático mais amplo e democrático para a participação política. Mais do que uma nova tecnologia, trata-se de um meio de comunicação, de interação e de organização social” (BRANDÃO JÚNIOR, 2005, p.3).

Castells (2006) define a internet como “mídia de massa individual”, pois a forma de acesso é diferenciada. O autor acredita que, se utilizada corretamente, a internet pode ser uma importante ferramenta política nas mãos da sociedade, podendo tornar o desenvolvimento econômico e social mais justo, “desenvolver a nossa responsabilidade individual, como seres humanos informados e conscientes dos nossos deveres e direitos, com confiança nos nossos projetos” (2004, p.324).

Já a importância da internet como fonte de informação para os políticos é reforçada por alguns autores (COUTINHO, 2010). Se devidamente rastreadas, as redes sociais podem se tornar uma importante fonte de informação para os candidatos utilizarem em suas campanhas políticas.

[...] é possível que os movimentos de opinião observáveis nas redes sociais antecipem tendências que somente serão captadas posteriormente (e com um custo muito mais elevado) através das pesquisas tradicionais de intenção de voto e temas de campanha. Se e como os partidos estarão atentos a isso é outra história, mas sem dúvida o surgimento de fontes “alternativas” de informação sobre intenção de voto e desempenho de candidaturas vai tornar mais difícil a vida dos institutos, e mais complexa a discussão sobre a estruturação da Esfera Pública no século XXI (COUTINHO, 2010, p.18).

Nesse cenário com sua dinamicidade e mensurabilidade, a internet possibilita aos estrategistas de marketing das campanhas eleitorais a fazer ajustes em tempo real conforme as avaliações de respostas postadas nas mídias sociais.

O tracking (pesquisa diária) mensura mudanças súbitas no humor do eleitorado e serve para incorporar fatos novos à campanha, fazer pré-teste de peças de comunicação que possam ser avaliadas quantitativamente, como conceitos e slogans e, benchmark para acompanhar a concorrência (AZEVEDO JUNIOR *et al*, 2015). No entanto, Azevedo Júnior *et al*. (2015) chamam a atenção para que esse tipo de acompanhamento da opinião pública deve considerar que os resultados apresentam viés em virtude de sua composição não ser representativa de toda a sociedade, as classes mais baixas ainda estão fora das redes sociais. Conforme dados do

IBGE coletados pelo Critério Brasil 2015 (ABEP, 2015) no Brasil ainda há baixa penetração nas classes D e E, que representam aproximadamente 26% da população brasileira (54 milhões de pessoas).

5.3.4 Cenário das redes sociais nas eleições brasileiras de 2014

Segundo o Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), foram 107,7 milhões de internautas no país em 2014 contra 99,2 milhões em 2013, com previsão de 125,9 milhões em 2018 (último ano da análise do eMarketer). A proporção de domicílios brasileiros com computador passou de 25% em 2008 para 49% em 2013. Mas ainda há muita gente sem acesso à rede: com base em dados da Pesquisa Nacional por Amostra de Domicílios (Pnad), do IBGE, 24,2 milhões de lares de renda de até 2 salários mínimos (em torno de R\$ 1,4 mil) não estão conectados à Internet. O mesmo vale para 7,5 milhões de lares na área rural do país.

O uso massivo das redes sociais pelos usuários brasileiros principalmente em relação à política eleitoral foi comprovada por pesquisas. A pesquisa Datafolha⁸⁰, realizada em outubro de 2014 com 4.389 pessoas em 257 municípios brasileiros mostrou que 75% dos internautas inscritos em redes sociais costumam ler notícias sobre as eleições por meio da sua página. Dos entrevistados 46% compartilham notícias sobre eleições em suas redes (Figura 9).

Figura 9: Eleitores conectados às redes sociais em 2014

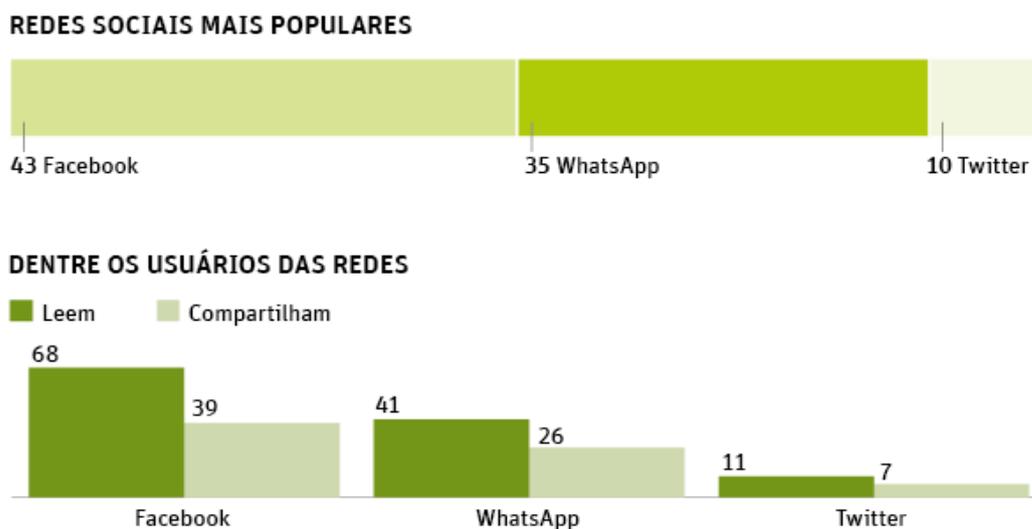


Fonte: Datafolha (2014)

⁸⁰ A pesquisa completa pode ser vista no site da Folha de São Paulo. Disponível em: <http://www1.folha.uol.com.br/poder/2014/10/1535693-nas-redes-sociais-75-leem-e-46-compartilham-sobre-eleicao-diz-datafolha.shtml>. Acesso em: 18 Jan. 2015.

Das redes sociais (Figura 10), o Facebook é a mais popular entre os usuários (43%), seguida do WhatsApp (35%) e em terceiro está o Twitter (10%). A pesquisa também observou que no universo de usuários a maioria usa o Facebook (68%) para saber sobre as eleições e os candidatos. O WhatsApp é também apontado por 41% e o Twitter por 11%.

Figura 10: Redes Sociais mais populares



Fonte: Datafolha (2014)

O levantamento Datafolha também mostra que de cada 10 eleitores, 6 têm acesso a internet. Entre os mais jovens – 16 a 24 anos – o acesso é de 9 a cada 10; entre os que têm entre 25 e 34 anos, de 8 de cada 10. Portanto, os jovens lideram mais esses espaços de interação e informação.

A pesquisa também levou em consideração o poder da internet como instrumento de convencimento político. Para 19% dos entrevistados, as redes sociais influenciaram muito na escolha do voto no primeiro turno; outros 20% dizem que influenciaram um pouco.

O hábito de declarar o voto também foi pesquisado pelo Datafolha. Dos entrevistados 20% dizem que anunciaram o candidato preferido no primeiro turno e 22% afirmaram que até o presente momento tinham anunciado o seu candidato nas suas redes. É preciso considerar que sobre esse item da pesquisa o Datafolha realizou a pesquisa uma semana antes do segundo turno das eleições, não conseguindo ter uma dimensão mais realista da campanha de segundo turno a respeito desta pergunta.

Nas eleições de 2014 a corrida eleitoral, principalmente as disputas presidenciais, entraram para a história das campanhas eleitorais. Os ataques verbais dos debates (entre

candidatos e entre eleitores) movimentaram e colocaram as mídias sociais em evidência na política brasileira. No Facebook, por exemplo, os brasileiros promoveram o maior número de interações já visto em outras disputas. Segundo pesquisas, foram contabilizados 674 milhões de interações, incluindo postagens, compartilhamentos, likes e comentários, entre 6 de julho e 26 de outubro. No Twitter, no mesmo período, foram 40 milhões de mensagens. E pela primeira vez foi usado o WhatsApp para enviar mensagens privadas por onde se espalharam tantas conversas de eleitores quanto peças de propaganda de candidatos. Ao contrário do Facebook e Twitter, que divulgam dados sobre a atividade de seus usuários, o WhatsApp guarda o anonimato.

No que diz respeito ao seu uso, não é possível quantificar o uso do aplicativo durante o período eleitoral, ou mesmo fora dele. Mas os números mostram como é grande o número de usuários que estão aderindo ao serviço. Em abril de 2014 eram 45 milhões de brasileiros, já em agosto do mesmo ano o número foi de 600 milhões⁸¹. Para especialistas o WhatsApp se configura como uma nova força nas estratégias de comunicação direta com os eleitores. Laura Olin, responsável pela campanha digital de Barack Obama em 2012, afirmou em uma reportagem para um veículo brasileiro⁸², que o uso do App será amplamente usado nas eleições norte-americanas de 2016.

Diante desse cenário, vemos cada vez mais as campanhas eleitorais focadas em buscar dados e informações de diversos ambientes para compor as suas estratégias de ação eleitoral.

5.4 CAMPANHAS ELEITORAIS CENTRADAS EM DADOS

Nesta parte do trabalho busca-se apresentar discussões em torno de campanhas eleitorais atuais que são desenvolvidas com base em análises de dados e informações sobre os cidadãos e que tem por objetivo informar e orientar os esforços táticos e contribuir com as estratégias eleitorais, principalmente na comunicação com os eleitores. Para o jornalista Sasha Issenberg⁸³ (2013), essa preocupação é tratada como uma nova ciência do voto adotada em campanhas eleitorais.

Nos itens anteriores discutiu-se a importância que as pesquisas eleitorais passaram a ter nos processos de disputas do voto. Para influenciar os eleitores, saber quais deles conquistar e se posicionar melhor perante esse público-alvo é necessário ter um conhecimento

⁸¹ Disponível em: <http://veja.abril.com.br/noticia/vida-digital/eleicoes-2014-revelam-novo-cabo-eleitoral-o-whatsapp>. Acesso em: 16 Jan 2015.

⁸² Declarações foram dadas para a Revista Veja e que podem ser encontradas na reportagem intitulada Eleições 2014 revelam novo cabo eleitoral o WhatsApp, por Claudia Tozetto, publicada em 01 de novembro de 2014.

⁸³ Sasha Issenberg deu entrevistas para a autora da tese durante a elaboração das pesquisas.

prévio do comportamento dos eleitores. Dessa forma, se faz necessário o emprego de métodos mais sistemáticos e científicos para conhecer o mercado eleitoral.

Nesse contexto, encontram-se o processo de segmentação eleitoral, que a partir da micro-segmentação dos eleitores chamado de *microtargeting* contribui para a tomada de decisões da campanha eleitoral. O *microtargeting* é muito parecido com a segmentação demográfica e geográfica tradicional, a diferença está que ao segmentar os nichos, podem-se trabalhar características mais precisas dos indivíduos a partir de cruzamentos de diferentes dados vindos em formatos e fontes diferenciadas.

Parte-se do pressuposto de que o emprego do *big data* na busca por características e individualidades comuns de eleitores possibilitadas com as técnicas e metodologias de microsegmentação em campanhas eleitorais têm se constituído num importante campo de investigação, mas pouco explorado tanto do ponto de vista do mercado eleitoral quanto das pesquisas científicas.

Para os profissionais que já fizeram usos do conceito, principalmente das análises de dados e informações, o *big data* se mostra como uma ferramenta importante em estratégias e ações de campanha. Campanhas antes centravam suas estratégias de comunicação com base apenas em pesquisas políticas tradicionais, mas com a possibilidade da Internet e das mídias sociais digitais, isso teve que ser reformulado. Especialistas da área de ciência política têm ressaltado a importância do uso do *big data* em campanhas eleitorais. Durante uma palestra, o cientista político e professor Adriano Oliveira (2013)⁸⁴ observou que iniciativas inovadoras na área de *big data* podem ser sentidas na política eleitoral.

As iniciativas inovadoras na área de analíticas envolvendo *big data* têm provocado rupturas na forma de realizar processos que envolvem análise de dados caracterizados por grande volume, enorme variedade na forma de acesso, armazenamento, e integração com fontes de dados relacionadas e pela alta velocidade com que são produzidos, modificados e compartilhados. (OLIVEIRA, 2013)

Para o cientista político as campanhas políticas precisam começar a repensar suas estratégias em relação às pesquisas de opinião, principalmente olhando as mídias sociais como potenciais espaços de coleta de dados e informações de eleitores.

No Brasil pouco ainda se faz referências sobre o emprego do *big data* em estratégias de campanhas eleitorais. Iniciativas foram citadas nas campanhas de 2010 e 2012 no Rio de

⁸⁴ A palestra foi realizada no dia 04 Nov. 2013, em Recife no V Seminário de Ciência Política, promovido pelo Centro Universitário Maurício de Nassau (UNINASSAU).

Janeiro, nas campanhas de Sérgio Cabral para governo do Estado e Eduardo Paes na campanha para a prefeitura (MOURA, 2011).

Aqui estabelecem discussões sobre essa temática a partir dos estudos feitos por Nickerson e Rogers (2014) e das considerações de Sasha Issenberg (2013) e Maurício Moura (2011). Para ilustrar essas discussões são apresentadas as experiências do uso de *big data* em campanhas americanas, como a do presidente Barack Obama, na disputa eleitoral de 2012. A apresentação dessa experiência se baseou em reportagens, artigos, sites, blogs e algumas bibliografias que foram produzidas até hoje sobre essa temática. Além disso, foi realizada uma entrevista com o jornalista americano Sasha Issenberg, que estuda campanhas eleitorais que usam de análises de grandes volumes de dados e informações em suas estratégias na conquista de votos.

5.4.1 O eleitor é o alvo

O objetivo geral de campanhas políticas é maximizar a probabilidade da vitória do candidato. Com essa finalidade, todas as facetas de uma campanha são avaliadas pelo número de votos que essa atividade irá gerar e qual o custo benefício. Esse pensamento é baseado num planejamento racional da disputa eleitoral. É preciso lembrar que as campanhas a partir de 1990 passaram a ser mais profissionalizadas e, portanto, também mais caras e contam cada vez mais com recursos escassos para serem gastos em suas estratégias (NORRIS, 2000).

Segundo Nickerson e Rogers (2014), para executar uma análise de custo-benefício, as campanhas precisam de previsões precisas sobre as preferências dos eleitores, seus comportamentos esperados e as suas respostas em relação à campanha. Para esses autores, os esforços para aumentar o número de eleitores são contraproducentes se a campanha mobiliza as pessoas que apoiam o adversário (NICKERSON; ROGERS, 2014, p. 51).

No cerne dessa discussão, está a segmentação do eleitorado, que coloca o eleitor de forma individualizada como centro das ações de marketing e comunicação numa campanha eleitoral. Diante da disponibilidade de grandes volumes de dados sobre os indivíduos, proporcionada pelas análises de *big data* (já apresentadas no decorrer da tese), as campanhas passam a operar com a lógica racional e matemática, buscando assim um custo/benefícios nas suas ações de comunicação. Segundo Tourin (2013), o *big data* possibilita uma segmentação bem mais precisa da população. Com o *big data* as fontes de informação se ampliam consideravelmente, podendo chegar ao indivíduo.

A prática da segmentação no contexto da política eleitoral é tão antiga quanto as eleições. A segmentação eleitoral é proveniente do marketing comercial. O conceito de

segmentação evoluiu de um conceito econômico até se tornar, atualmente, em uma estratégia de conquista de mercados. A expressão “Segmentação de Mercado” foi cunhada por Wendell Smith, em 1956, em seu artigo “Product Differentiation and Market Segmentation as Alternative Marketing Strategies”. Apesar de ter sido essa a primeira vez que se usou esse termo com esta aplicação, a primeira referência ao conceito foi feita por J. Frederick, em 1934, em seu livro sobre Marketing Industrial, onde já se referia à sua utilização por algumas empresas.

Segundo Kotler (1996, p.257), "segmentação de mercado é o ato de dividir um mercado em grupos distintos de compradores com diferentes necessidades e respostas." Lobos (1993, p.269) compartilha da definição de Kotler e acrescenta: "divisão de mercado em populações de ‘clientes’ que apresentam certas ‘características’. É impossível prestar serviços igualmente excelentes a todos os clientes imagináveis; daí a necessidade de segmentar”. Para McCarthy e Perreault (1997), a segmentação é colocada como um processo de agregação de pessoas com necessidades similares. Então, a segmentação é um processo que tira do todo heterogêneo e subdivide-se em mercados menores que tenham características semelhantes. Variáveis são usadas para segmentar mercados consumidores e estas são classificadas em grupos de consumidores. Dependendo, o autor nos traz um número diferente de variáveis. Kotler (1996, p.241) divide as principais variáveis em:

- **geográfica** (região, clima etc.);
- **demográfica** (idade, sexo, renda, religião, raça, nacionalidade etc.);
- **psicográfica** (classe social, estilo de vida, personalidade);
- **comportamental** (benefícios, atitudes, grau de lealdade etc.).

McCarthy e Perreault (1997, p.73) estabelecem as variáveis em número menor e as divide em três: **comportamental, geográfica e demográfica**. Já Westwood (1997, p.72) subdivide em um número diferente dos dois autores acima, estabelecendo em cinco o número de variáveis: **classe social, demográfica, geográfica, geodemográfica, psicográfica**. Todos os autores citados concordam com a necessidade de classificar os usuários finais, e que cada grupo destes comporta uma estratégia mercadológica bem definida. Em qualquer que seja o conjunto de variáveis escolhidas para conduzir a campanha eleitoral, é nítido que o alvo aqui é o eleitor.

A segmentação de eleitores segue as mesmas premissas das adotadas no marketing comercial. Segundo os especialistas, a segmentação do eleitorado, além ser uma estratégia

necessária, torna a campanha mais barata. A importância da segmentação torna-se mais evidente em cidades de médio e grande porte, como São Paulo e Rio de Janeiro por exemplo, em que determinados bairros são maiores que muitas cidades brasileiras precisam adotar esse tipo de estratégias em campanhas eleitorais.

A importância da segmentação nos mercados político é enfatizada por Hooley e Saunders (1993) em que as sofisticadas campanhas políticas podem ser elaboradas através das variáveis geográficas, demográficas e nas variáveis que intensificam o foco na demanda e nos fatores de preferência, como segmentos com hábitos semelhantes dos meios de comunicação, atitudes e características ou outro estilo de vida ou benefício. A essência do posicionamento consiste em aceitar as percepções como realidade e então reestruturar essas percepções a fim de criar a posição desejada, como bem esclarecem Ries e Trout (2002), quando afirmam que posicionamento “é o modo como você se diferencia na mente de seu cliente potencial”.

Assim, o posicionamento tem como ênfase as características distintivas que fazem um candidato diferente de seus competidores e atraente para seu público, como o ato de relacionar uma faceta de suas ideias a um conjunto de expectativas, necessidades e desejos de eleitores. Portanto, posicionamento é um conceito crucial, que lembra que as escolhas dos eleitores são feitas numa base comparativa, e um candidato só será escolhido se fizer parte claramente de um processo de seleção.

Na segmentação, o eleitorado pode ser dividido em três grupos:

- (1) **Os eleitores que certamente votarão no candidato.** Este segmento requer menos custos, ou seja, tempo e recursos. São os eleitores que já conhecem o candidato e não o rejeitam. Nesse caso, a chance do voto é bem maior aqui.
- (2) **Os eleitores que estão determinados a não votar no candidato.** O tempo e o custo para persuadir esse eleitor são inviáveis, levando em conta a pequena probabilidade de sucesso. São os eleitores que rejeitam determinada candidatura e dificilmente mudarão de opinião. Nesse caso, é melhor a campanha não ser direcionada a esses eleitores.
- (3) **Em maior segmento, estão os eleitores que podem vir a votar no candidato.** É justamente aqui que deve se concentrar o maior esforço da campanha. O “*target*”, ou eleitorado alvo, é aquele que quantitativamente pode eleger o candidato, já que o primeiro segmento pode não ser muito expressivo. É nesse público que a campanha concentra suas atenções e investimentos.

No entanto, é necessário segmentar ainda mais o eleitorado para compreender melhor a quantidade e localização. Nesse sentido, o *microtargeting* se configura numa técnica mais indicada para dar esse grau de detalhamento do eleitorado a ser atingido em uma campanha. O uso da grande massa de dados dos eleitores, de diferentes formatos e fontes configura-se no *big data*, já tratado de forma mais aprofundada anteriormente nesse trabalho, e que contribui para o grau de detalhamento do eleitorado que a campanha procura atingir.

Para alguns estudiosos (NICKERSON; ROGERS, 2014), embora a adoção dos novos métodos analíticos não transformarem radicalmente a forma das campanhas de operar, o seu uso mostra algumas vantagens competitivas em relação àquelas que não a usam. Esse método de fazer campanhas eleitorais levou os partidos políticos a uma corrida armamentista para buscar um crescente volume de dados para ajudar na conquista de votos (NICKERSON; ROGERS, 2014, p. 51). Essa é uma realidade dos partidos americanos, que desde 2000 intensificaram suas atividades na busca cada vez mais intensa por dados e informações dos cidadãos americanos. Como por exemplo⁸⁵, para as próximas eleições de 2016, o partido Democrata já iniciou a mobilização junto aos eleitores com ações claras, para não deixar que um Republicano seja o próximo presidente, buscando fortalecer a campanha de Hillary Clinton para presidente dos Estados Unidos. Entre as estratégias de marketing político está a criação de uma base de dados atualizada de contatos, que será usado futuramente na corrida eleitoral de 2016. Os eleitores terão que deixar a informação pessoal, o que permitirá futuramente à máquina partidária realizar campanhas altamente direcionadas ao seu eleitorado.

Nos seus estudos, Nicherson e Rogers (2014) destacam por que as campanhas políticas precisam de dados. Segundo os autores, as campanhas contemporâneas precisam de dados para manter suas listas de cidadãos, para que possam entrar em contato durante o processo eleitoral, sejam para pedir votos ou doações de eleitores (no caso americano). As informações devem ser precisas e devem vir dos cidadãos, voluntários e doadores. Esses dados servem para construir modelos preditivos, que têm como objetivo fazer segmentação da campanha, comunicações mais eficientes e para apoiar estratégias mais amplas (NICKERSON; ROGERS, 2014). Esses autores estabelecem os modelos preditivos em três categorias que são chamados de “escores preditivos”. São eles: escores de comportamento, escores de apoio e escores de respostas que são descritos a seguir.

⁸⁵ Informações retiradas do texto “Stop: como os partidos americanos usam os princípios da persuasão para mobilizar eleitores”. Disponível em: <https://dissonanciacognitiva.wordpress.com/2015/04/15/stop-como-os-partidos-americanos-usam-os-principios-da-persuasao-para-mobilizar-eleitores/>

Escores de Comportamento

É o comportamento passado e informações demográficas para calcular de forma explícita a probabilidade de os cidadãos poderem participar em determinadas atividades políticas (comparecimento às urnas; doações; voluntariado). Nesse caso, as informações sobre o comportamento do voto, que ficam registrados em arquivos oficiais eleitorais, que são de uso público, podem dizer muito do eleitor e sua história com o voto e as eleições. Por exemplo, se votou antecipadamente, como votou, se compareceu às urnas ou mesmo se não compareceu. Assim, é possível identificar no grupo de eleitores se a pessoa tem alta propensão de votar ou não (NICKERSON; ROGERS, 2014, p.9 -10). “A localização geográfica das residências dos cidadãos também pode fornecer informações valiosas, porque as campanhas podem mesclar dados relevantes do censo com informações sobre os eleitores” (NICKERSON; ROGERS, 2014, p.10). Dados do Censo, tais como renda média familiar, o nível médio de educação, número médio de filhos, e por agregado familiar, étnico-distribuição são úteis para o desenvolvimento de uma série de escores preditivos.

Escores de apoio

Têm por objetivo prever as preferências políticas dos cidadãos. No mundo ideal da campanha, consultores entrariam em contato com todos os cidadãos e perguntariam sobre seu candidato e suas preferências. No entanto, no mundo real de restrições orçamentárias, as campanhas entram em contato com um subconjunto dos cidadãos e utilizam as suas respostas como dados para desenvolver modelos que predizem as preferências do resto dos cidadãos que estão registrados para votar. Os escores de apoio variam tipicamente **0-100** e geralmente são interpretados da seguinte forma: se numa amostra de 100 cidadãos com uma pontuação de X, X por cento preferem o candidato. A pontuação de apoio de "0" significa que ninguém em uma amostra de 100 cidadãos apoiaria o candidato, "100" significa que todos na amostra apoiariam o candidato, e "50" significa que metade da amostra apoiaria o candidato. Essas pontuações suportam apenas prever as preferências no nível do agregado, e não a nível individual. Isto é, as pessoas com contagens de apoio de 50 não são necessariamente indecisos ou que não votariam no candidato e, de fato, pode ter fortes preferências. Mas quando os cidadãos têm pontuações de suporte 50, isto significa que é difícil prever o suas preferências políticas.

Escores de respostas

São aqueles que apresentam previsões de como cidadãos irá responder a campanha de divulgação. Segundo Nickerson e Rogers (2014) embora existam justificativas teóricas a

respeito de quem pode ser mais sensível a agrados para votar e tentativas de persuasão em geral, prever quais indivíduos serão mais e menos sensível a determinadas comunicações diretas em um determinado contexto eleitoral é difícil. As campanhas podem usar experimentos de campo totalmente ao acaso, para medir a resposta a uma tática de campanha (escores preditivos) que guia a segmentação de decisões (ISSENBERG 2012A, 2012b, 2012c).

Nickerson e Rogers (2014) e Moura (2015) compartilham da ideia de que a maioria da informação útil do eleitor deve ser colhida diretamente com os próprios indivíduos. Fazendo referências às campanhas eleitorais americanas, Nickerson e Rogers dizem que aqueles que doaram e foram voluntários no passado são as perspectivas para recrutamento no futuro. Além disso, os atributos desses indivíduos podem ser usados para o desenvolvimento de escores de comportamento para identificar outros que possam estar propensos a doar ou a ser voluntários na campanha. Do mesmo modo, informações sobre indivíduos que responderam ao telefonema da equipe ou porta a porta no passado podem ser usadas para desenvolver análise de comportamento em campanhas eleitorais futuras. Os dados de pessoas que se inscrevem para receber e-mails em campanhas também podem ser úteis. É possível fazer um monitoramento das suas ações em relação à conta cadastrada.

O valor do uso de escores preditivos para o direcionamento da campanha tornou-se amplamente reconhecido por campanhas eleitorais durante os últimos cinco anos nos Estados Unidos e também foi sistematicamente utilizado na campanha de reeleição de Barack Obama em 2012. Os escores preditivos são usados para atingir todos os aspectos da campanha de divulgação como as ações porta-a-porta, mala-direta, telefonemas, e-mail, colocação de anúncio de televisão, mídias sociais (NICKERSON; ROGERS, 2014, p.19)

5.4.2 Microtargeting eleitoral

Para Moura (2011) um dos principais desafios dos estrategistas brasileiros é segmentar o eleitorado. “Diferente dos Estados Unidos, o Brasil não autoriza o acesso aos dados dos cartões de crédito e os dados públicos, apesar dos avanços na qualidade do trabalho do IBGE, ainda não são suficientes para atingir o eleitor em sua residência” (MOURA, 2011, p. 58).

O *microtargeting*, largamente usado em campanhas eleitorais nos Estados Unidos, principalmente desde as disputas políticas de 2000, a partir da construção de bancos de dados

sobre os cidadãos permite aos políticos conhecer quem é cada eleitor, como e quando toma as suas decisões, que temas prefere, se está disposto a absorver uma comunicação mais ousada⁸⁶.

A técnica para os especialistas tem se mostrado atraente para dar vantagem competitiva aos candidatos que disputam uma eleição principalmente considerada acirrada. O *microtargeting* reúne dados sobre os que apoiam o candidato juntamente com os dados de consumo, informações públicas dos eleitores e as atividades nas mídias sociais. Trata-se de uma nova ferramenta para os candidatos e campanhas que responde às suas questões mais fundamentais como: **Quem apoia meu candidato? Onde posso encontrá-lo? Como posso convencer os outros para apoiar o meu candidato?**

Segundo Lundry (2013), quando bem feito, o *microtargeting* responde a cada uma dessas questões e faz isso de forma que a segmentação tradicional não pode fazer. Para alguns especialistas, o *microtargeting* nada mais é do que uma metodologia renomada da antiga campanha feita porta-a-porta, com um diferencial, hoje com a tecnologia e com o *big data* é possível dinamizar essas ações de comunicação direta. O *microtargeting* ficou mundialmente conhecido quando o presidente americano Barack Obama na sua campanha de 2012 organizou uma estrutura tecnológica para saber quem procurar e o que dizer e que será detalhada adiante neste trabalho.

Sasha Issenberg (2015) observa que o *microtargeting* permite às campanhas falarem diretamente aos eleitores exatos, que os candidatos precisam atingir para vencer.

“Eles (as campanhas) terão uma previsão da probabilidade de que você é proprietário de uma arma. Isso é, obviamente, uma questão muito sensível. Você não quer falhar se você é um candidato e envie algo sobre o controle de armas para alguém que você pensa possuir uma arma e vice-versa. Assim, a campanha pode ter certeza que quando eles estão enviando algo sobre o controle de armas é apenas para alguém que está muito certo é dono de uma arma” (ISSENBERG, 2015).

Já no contexto brasileiro, a metodologia foi trazida pelo pesquisador Maurício Moura a partir da campanha de 2010, ao trabalhar a campanha de Sérgio Cabral (PMDB) que disputava a reeleição para governo do Estado do Rio de Janeiro. Moura relata em um artigo (2011), que a partir de uma série de pesquisas, através de coleta de dados e informações de eleitores fluminenses foi possível observar a existência de uma fatia muito grande de indecisos entre as mulheres com mais de 50 anos, no interior do Estado. O passo seguinte da metodologia foi dirigir mensagens especificamente a esse público. Para chegar a essa conclusão o pesquisador usou a metodologia da microsegmentação, técnicas matemáticas e

⁸⁶ Informações disponível em: http://translate.google.com.br/translate?hl=pt-BR&sl=en&u=http://www.realclearpolitics.com/articles/2014/12/08/all_politics_is_customizable_the_pitfalls_of_micro-targeting_124856.html&prev=search. Acesso em: 28 Jan 2015.

modelos preditivos. Na campanha de reeleição para prefeito do Rio de Janeiro em 2012 do candidato Eduardo Paes (PMDB) *microtargeting* também foi amplamente utilizado pelo pesquisador Maurício Moura.

5.5 CASE BARACK OBAMA 2012

Essa parte do trabalho é dedicada a descrever como a campanha política de 2012 do democrata Barack Obama que concorreu à reeleição à Casa Branca nos Estados Unidos em 2012, centralizou sua campanha com o uso de análises de grandes massas de dados e técnicas de *microtargeting*. A campanha foi considerada como um exemplo emblemático de como *big data* ajudou na vitória do candidato⁸⁷ (MORAES, 2013).

Para desenvolver essa parte do trabalho foram feitas pesquisas em diversos documentos e reportagens de revistas nacionais e internacionais e a partir de depoimentos do jornalista americano Sasha Issenberg⁸⁸, que é especialista em coberturas de campanhas eleitorais americanas.

Antes, vale ressaltar que no sistema eleitoral americano o voto do eleitor não é obrigatório e a campanha política é disputada num longo período de quase um ano. As propagandas políticas na televisão são pagas e a maior parte do dinheiro das campanhas políticas vem de arrecadações de eleitores⁸⁹. Nos EUA, não ganha eleição quem obtém mais votos populares, mas sim quem tem mais votos no Colégio Eleitoral, composto pelos 538 eleitores repartidos por cada Estado. Os partidos democrata e republicano se preocupam com a coleta de informações sobre os eleitores a partir do comportamento de consumo desde 2000 (ISSENBERG, 2015, 2013; MOURA, 2015).

⁸⁷Big data ajudou Obama a ganhar eleições. Disponível em: < http://info.abril.com.br/noticias/ti/big-data-ajudou-obama-a-ganhar-eleicoes-15012013-25.shl?utm_source=messiaslg&utm_medium=linkedin&utm_campaign=Feed%3A+NoticiasINFO-TI+%28Not%C3%ADcias+INFO++TI%29>. Acesso em: 20 Mar. 2013.

⁸⁸ Sasha Issenberg deu entrevista por e-mail para esta tese.

⁸⁹ O financiamento de campanhas políticas para eleições federais (Presidência, Senado ou Câmara dos Representantes) nos Estados Unidos é supervisionado pela Federal Election Commission (Comissão Eleitoral Federal, ou FEC, na sigla em inglês), uma agência federal independente. A maior parte do financiamento vem de fontes privadas — que podem ser pequenos doadores individuais (pessoas que contribuem com US\$ 200 ou menos), grandes doadores individuais (que contribuem com mais de US\$ 200), comitês de ação política (os chamados PACs) e grupos cívicos — ou mesmo de autofinanciamento, nos casos em que o candidato financia a campanha com seu próprio dinheiro. A legislação estabelece limites para as doações. Indivíduos podem doar até US\$ 2,5 mil diretamente a um candidato ou US\$ 30,8 mil a um comitê nacional de partido político. Empresas e sindicatos são proibidos de fazer doações diretamente a candidatos ou partidos políticos, mas podem doar dinheiro aos PACs. (CORRÊA, 2013). Informações retiradas da reportagem publicada disponível em: http://www.bbc.co.uk/portuguese/noticias/2013/07/130710_financiamento_eleicoes_dg. Acesso em: 10 Mar. 2014.

Entre as estratégias de marketing usadas na campanha Obama for America (OFA) destacam-se as de segmentação de mercado e de relacionamento com eleitores, o chamado marketing *one to one* (um a um). Obama conduzia todas as informações que vinham das análises de *big data* para dar o direcionamento das suas ações. Uma das coisas que ajudou a campanha foi a definição detalhada do mapa eleitoral das regiões americanas, que permitia à equipe de campanha focar onde se achava que podia ganhar. Dessa forma, os democratas conseguiram ganhar em estados como Carolina do Norte (mobilizando jovens, negros, mulheres solteiras), Louisiana e Indiana – onde não venciam desde Lyndon Johnson, em 1964.

Segundo Issenberg (2013) as estratégias de analítica de dados e de segmentação da campanha de Obama de 2012 começaram a ser desenvolvidas cinco anos antes pelo partido dos democratas (ISSENBERG, 2013). Foram realizados cadastros de pessoas favoráveis às suas ideias e seus candidatos. Nesse período foram coletados 13 milhões de endereços de e-mails de simpatizantes. Destes, três milhões efetuaram doações. Além disso, Issenberg ressalta que o ponto de partida para a conquista de eleitores de forma mais individualizada vem de outras campanhas do democrata (ISSENBERG, 2013).

Barack Obama disputou duas eleições para a presidência dos Estados Unidos. Uma em 2008, contra o republicano John McCain e outra em 2012 contra o também republicano Mitt Romney. Nas suas duas eleições o democrata saiu vitorioso. Na visão dos especialistas a eleição de 2008 representou para Obama e outros democratas um crescente esforço de base no mundo digital. No dia 31 de abril de 2006, o jornal The New York Times publicou uma reportagem que apontava para uma intensa transformação da política norte-americana por causa da Internet. Na medida em que se aproximavam as eleições presidenciais para a Casa Branca, o jornal fazia um diagnóstico dizendo que seriam reescritas as regras políticas de propaganda, arrecadação de fundos, mobilização de apoiadores e mesmo da disseminação de informações negativas. A reportagem dizia⁹⁰:

Democratas e republicanos estão ampliando fortemente o uso do , websites interativos, blogs de candidatos e partidos, mensagens de texto para arrecadação de dinheiro. A internet, dizem eles, parece ser muito mais eficiente e com menos custo do que as ferramentas tradicionais da política⁹¹ (tradução nossa).(NAGOURNEY, 2006).

⁹⁰ Escrevia o jornalista Adam Nagourney numa reportagem “Politics Faces Sweeping Change via the Web” Disponível em: www.nytimes.com/2006/04/02/washington/02campaign.html?pagewanted=all&_r=0. Acesso em: 20 Mar. 2014.

⁹¹ Democrats and Republicans are sharply increasing their use of , interactive Web sites, candidate and party blogs, and text-messaging to raise money, organize get-out-the-vote efforts and assemble crowds for a rallies. The Internet, they said, appears to be far more efficient, and less costly, than the traditional tools of politics (Texto original da citação).

O resultado e os impactos da campanha de 2008 ficaram marcados como um importante case no marketing político contemporâneo. O modelo de campanha que levou Obama à presidência dos Estados Unidos foi estudado por especialistas como Umair Haque, do Havard Media Lab. Em artigo publicado pelo site *Harvard Business Publishing*, ele disse: “A estrutura de campanha Obama está para as campanhas tradicionais como a empresa Google está para uma empresa convencional”. Em um país, como os Estados Unidos, onde o voto não é obrigatório, o fator de engajamento que é propriedade das redes sociais fez grande diferença para a eleição de Barack Obama⁹². Na opinião, Oliveira (2014) da mesma forma em 2012, os usos de dados e também das mídias sociais digitais contribuíram para a sua vitória à Casa Branca e tornaram a campanha política como referência no uso de *big data* como estratégia eleitoral.

Jim Messina, gerente de campanha de Obama em 2012, afirmava na época em entrevistas para os meios de comunicação americanos que iriam medir tudo na campanha eleitoral de 2012. E foi o que a campanha OFA fez, todos os dados foram medidos e datificados. A campanha, que na primeira disputa em 2008 tinha sido definida por ideias de esperança e mudança, transformou-se em 2012 na campanha de dados (ISSENBERG, 2013). A campanha começou com uma base de dados de 10 TB (BigData-Startups, 2013) e teve um crescimento até o fim da disputa eleitoral de 50 TB (BURT, 2013).

Os objetivos da campanha de Obama giravam em torno do convencimento de pessoas certas a fazerem doações e buscava apoios à campanha, bem como o aumento do comparecimento às urnas. Além disso, buscava manter a fidelidade do voto daqueles eleitores que já haviam votado em sua campanha em 2008.

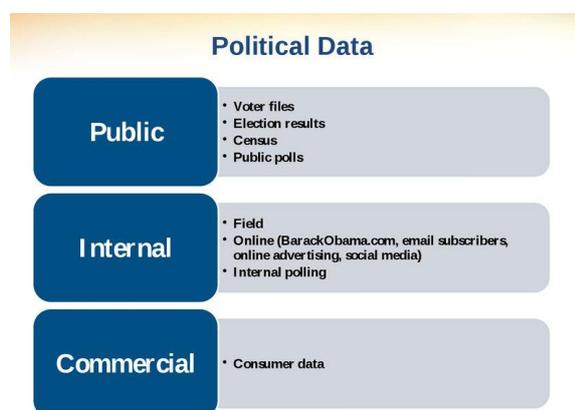
Entre as estratégias usadas para atingir esses objetivos, estavam: (1) Técnicas usadas por empresas para atrair consumidores, como marketing direcionado e microsegmentação; (2) Compra de dados em empresas como Experian e a Acxiom; (3) Enriquecimento de dados com informações de redes sociais (em especial Twitter e o Facebook); (4) Segmentação dos eleitores de acordo com o perfil de navegação no site oficial da campanha; (5) Criação de cluster de acordo com os gostos e características das pessoas (perfil político etc.); (6) Ações de comunicação individualizadas *one to one*.

⁹² Em 6 de novembro de 2012, Obama ganhou 332 votos no colégio eleitoral, superando os 270 votos necessários para que ele fosse reeleito como presidente. Com 51% de votos populares, Obama tornou-se o primeiro presidente democrata desde Franklin D. Roosevelt a ganhar duas vezes a maioria dos votos populares. (Informações retiradas do Wikipedia. Disponível em: < http://pt.wikipedia.org/wiki/Barack_Obama>. Acesso em: 3 Mar. 2014.

As informações vindas do documento oficial da campanha de Obama “*Inside the Cave*”⁹³ relata como eram feitas outras ações de coleta de informações e análises de dados na campanha. É importante destacar que a campanha de Obama em 2012 sabia o nome de cada um dos 69.456.897 americanos que votaram nele em 2008. Essas informações foram possíveis a partir do mapa eleitoral feito pela equipe de campanha. Esse mapa do eleitor americano começou a ser delineado com o então cientista político Peter Giangreco⁹⁴, considerado um dos maiores especialistas em mala direta dos Estados Unidos, que na campanha de 2008 foi responsável pela estratégia de marketing direto e *microtargeting* da campanha presidencial.

Toda a captação de dados na campanha de Obama for America (OFA) para a reeleição foi integrada. Foram reunidas o maior quantidade possível de dados de eleitores americanos, que vinham de diversas fontes tanto da web como fora dela (Figura 12). Para analisar os dados foram criados algoritmos e aplicações para lidar com esses dados coletados.

Figura 12 : Dados da campanha Obama



Fonte: CLASTER (2012)

Os dados vinham de campanhas passadas, empresas especializadas (como Experian e Acxiom), redes sociais, assinantes do aplicativo Obama for America⁹⁵. Um volume gigantesco de dados, chegando a 4 gigabytes processados por segundo, 8,5 bilhões de requisições ao banco de dados e 180 terrabytes de armazenamento em três data centers. O banco de dados da

⁹³ Documento disponível em <http://enga.ge/download/Inside%20the%20Cave.pdf>. Acesso em: 03 Marc. 2014. O documento também foi encaminhado para a autora deste trabalho por um dos integrantes da equipe da campanha via redes sociais Facebook.

⁹⁴ Peter Giangreco é formado em Ciência Política pela Universidade de Michigan e professor convidado das universidades de Chicago Loyola e Harvard.

⁹⁵ Página Oficial da campanha (<http://barackobama.tumblr.com/>).

campanha era abastecido com informações sobre os modos como as pessoas abriam o , tuítavam ou publicavam no Facebook e sobre mudanças demográficas.

Para os especialistas, a coleta e análise de dados foi algo nunca visto antes em campanhas anteriores. Essa base de dados extremamente detalhada foi o cérebro de toda a operação, mesmo da campanha fora da web. Ferramentas de análises como *Google Analytics*⁹⁶ foram fundamentais a partir de então. Ter acesso rápido e fácil a dados acionáveis foi essencial para a campanha de reeleição de Obama.

No trabalho de microsegmentação e análise desses dados, identificou-se na campanha que havia em Nova York um público interessado por um tipo de programa noturno peculiar, um possível jantar com celebridades, especialmente com a atriz Sara Jessica Parker e George Cloney. Nesse grupo verificou-se que tinham poder aquisitivo grande. A partir disso, a campanha lançou um concurso de doação para a campanha. Quem doasse poderia jantar com a atriz e o candidato Obama. A estratégia também foi aplicada com o ator George Cloney, em que mulheres de 40 a 42 anos manifestavam interesse em um jantar na companhia da celebridade. Em torno de 150 pessoas participaram do jantar na casa do ator George Cloney, cuja entrada custou 40 mil dólares. O evento foi organizado na residência do ator em Los Angeles, Califórnia (EUA) e Obama arrecadou com essa estratégia US\$ 15 milhões para sua campanha.

A campanha tinha por objetivo também valorizar o doador de quantias pequenas, de até U\$ 200 e que representavam 95% dos apoiadores (TIROBASCHI, 2013). Essa estratégia tinha como objetivo de inserí-lo na campanha. Para isso foram criadas experiências com sites e plataformas para aparelhos móveis como smartphones e tablets. “Criamos um sistema de “doação rápida” em que, para doar uma segunda vez, bastava um clique. Queríamos que as pessoas tivessem uma experiência o mais simples possível”, afirma Slaby para a Revista IstoÉ (2013). Com essas ações a campanha chegou a arrecadar quase US\$ 700 milhões de forma on-line, de 4 milhões de eleitores.

⁹⁶ *Google Analytics* é um serviço gratuito e é oferecido pela Google no qual, ao ativar-se o serviço por intermédio de uma conta do Google, e ao cadastrar-se um site recebe-se um código para ser inserido na página cadastrada e, a cada exibição, estatísticas de visitação são enviadas ao sistema e apresentadas ao dono do site. Foi criado principalmente como um auxílio aos webmasters para otimizar seus sites para campanhas de marketing. O *Analytics* é capaz de identificar além da tradicional taxa de exibição e hit de uma página, localização geográfica do visitante, forma com a qual chegou à página, sistema operacional, navegador etc. Embora muitas pessoas vejam o *Google Analytics* apenas como uma ferramenta de monitoramento de tráfego essa ferramenta na verdade funciona como uma poderosa ferramenta para tomadas de decisões relacionadas à internet.

5.5.2 Equipe e tecnologia

Para se ter a dimensão e importância que uma campanha eleitoral dá a determinados segmentos internos da sua campanha, leva-se em consideração o uso de tecnologias e métodos analíticos que são empregados na definição de suas estratégias. Isso, segundo especialistas, pode ser evidenciado pelo número de especialistas de cada área contratados pelas campanhas. Em relação a campanha de Obama, Oliveira (2014)⁹⁷ observa que:

O lado democrata dedicados à análise dos dados do perfil dos eleitores americanos com o objetivo de planejar e executar ações para persuadi-los a votar em favor do seu candidato era aproximadamente 10 vezes maior que o time republicano contratado para o mesmo tipo de trabalho. De acordo com os estudos e análises mencionados acima sobre o efeito big data na eleição, o uso eficaz de soluções analíticas implementadas por cientistas de dados e profissionais com perfil similar foi fator decisivo no resultado da eleição. As análises produzidas foram referências extremamente valiosas para orientar de forma assertiva as ações de marketing, o discurso e a estratégia de campanha em cada estado, que levaram à vitória do presidente Obama na disputa por um segundo período à frente do governo dos EUA (OLIVEIRA, 2014).

A campanha de Obama foi coordenada pelo especialista em dados Dan Wagner, que respondia como diretor de segmentação do Diretório Nacional do partido (DNC). Wagner tornou-se responsável pela coleta e análise de informações dos eleitores individuais por mala direta e telefone e esse trabalho começou em 2009. O especialista pediu ao departamento de tecnologia do DNC para desenvolver software que poderia transformar em tabelas, o que ele chamou de Gerente de Pesquisas (*Survey Manager*). Segundo Greenwald da Mit Technology Review, durante os 18 meses da campanha do presidente Obama, um grupo de análise de dados coletou e combinou milhares de dados sobre todo o público votante dos Estados Unidos.

Nate Lubin (2013), Diretor de Marketing Digital de Obama for America (OFA) diz que para manter as equipes digitais mais integradas e rápidas, era preciso fazer a comunicação coerente e abrangente disponível para todos em uma base quase em tempo real. "Ao longo da campanha, o *Google Analytics* nos ajudou a democratizar nossos dados na web, e permitiu que todas as nossas equipes tomassem decisões importantes rapidamente" (LUBIN, 2013). Lubin credita ao *Google Analytics* parte do sucesso da campanha de Obama. Segundo ele, a ferramenta foi essencial para que a equipe pudesse obter informações e tomar decisões que influenciaram o curso das estratégias da campanha.

Como mostra o documento oficial da campanha OFA *Inside the Cave* (2012) a campanha em 2012 tinha duas frentes de trabalho: uma digital comandada por Teddy Golf e

⁹⁷ Retirado do blog o povo: <<http://blog.opovo.com.br/bigdata/2014/05/02/big-data-nas-eleicoes-2014/>>

Joe Rospars, que tinham como função cuidar do público da campanha. Entre as atividades estavam: arrecadação de fundos, gerenciamento de e-mails e redes sociais. E uma tecnológica e analítica em departamentos separados, comandados pelo diretor de Inovação e Integração Michael Slaby. Sobre a campanha Michael Slaby⁹⁸ conta que:

[...] um dos mais interessantes foi uma plataforma chamada “dashboard”, uma maneira de conectar e unificar as atividades off-line e on-line de nossos voluntários. Todo o trabalho que cada um estava fazendo era registrado em um só lugar, como um escritório de campanha on-line. Os voluntários podiam se conectar, conhecer seus vizinhos que também faziam parte da campanha, saber quem eram os líderes locais e trabalhar juntos como um time. Também fizemos ferramentas on-line de registro de eleitores e manuais para ajudar nossos voluntários a usar o Facebook de maneira mais eficiente para atingir mais pessoas. [...] (Trecho da entrevista de Slaby concedida a revista IstoÉ em 11 de junho de 2013 para a jornalista Juliana Tirobaschi)

Slaby ficou conhecido como o “guru” digital de Obama, chefe de tecnologia da campanha que levou o então senador (Barack Obama) a seu primeiro mandato presidencial, em 2008. Em 2011, atuou como líder do departamento de inovação e integração da campanha de reeleição de Obama.

A campanha tinha como objetivo trazer outras pessoas especializadas para trabalhar na equipe. Assim em junho de 2011, os coordenadores da campanha começaram a busca por meio de suas listas de doadores, procurando pessoas com características e ocupações em tecnologia, dados ou design. A cada perfil encontrado eram disparados e-mails com a seguinte mensagem:

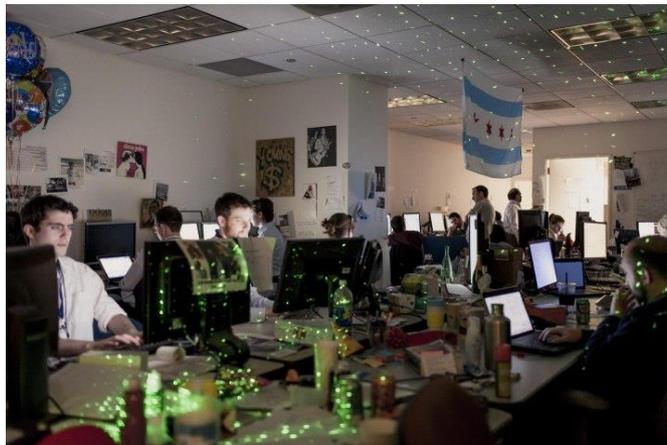
"You're one of very few people receiving this e-mail because, based on what you've told the Obama organization in the past. We think you might know someone who should quit his or her job and come work on the Obama campaign's digital team for the next 18 months."
"It won't pay very well. The hours are terrible Most people who come to work here will take a pay cut."⁹⁹

Em entrevista à Revista Veja (2012) um dos profissionais voluntários na campanha de Obama em 2012, o brasileiro Ricardo Cappra, explica como a equipe de dados trabalhava. Segundo Cappra, foi montado um centro de inteligência de dados chamado de The Cave (A Caverna), como demonstra a Imagem 2.

⁹⁸ Em entrevista a Revista Isto É no dia 11 de junho de 2013. Disponível em: <http://www.istoe.com.br/reportagens>. Acesso em: 20 Jun. 2014.

⁹⁹ Tradução livre: “Você é uma das poucas pessoas que receberam este porque, com base no que você disse a organização Obama no passado. Nós pensamos que você conhece alguém que pode parar seu trabalho e vir trabalhar na equipe digital da campanha de Obama nos próximos 18 meses”.
"Ele não vai pagar muito bem. As horas são terríveis A maioria das pessoas que vêm para trabalhar aqui vai ter um corte de pagamento ”.

Imagem 2 : O centro de operações do big data Obama



Fonte: Inside the Cave (2012)

Esse centro era que alimentava toda a estratégia de comunicação da campanha, onde ficaram 12 pessoas, que passaram 60 dias morando nesse local, isoladas, tendo acesso a todo tipo de dados e o que eles filtravam dali direcionava tudo à campanha. Cappra explica que na prática, essa equipe reportava para outra equipe de quatro pessoas – da qual ele (Cappra), dois americanos e um russo faziam parte – todas as informações mineradas a partir de dados. Por exemplo: eles entendiam o que estava acontecendo em uma região e sugeriam uma ação para os estrategistas da campanha. Tudo em tempo real. Assim foi possível, por exemplo, prever com 15 dias de antecedência que o desempenho de Obama no primeiro debate não seria satisfatório, permitindo que a equipe planejasse uma ação on-line, que foi ao ar uma hora depois do programa.

Todos os dados eram públicos e foram mensurados através de ferramentas *open source*: dados de redes sociais, órgãos públicos, partido etc. Com o estudo, foi possível também prever o resultado das eleições com 30 dias de antecedência. Foram mapeados todos os estados onde o Obama ganharia as eleições e de quanto seria o seu percentual de vitória (CAPPRA, 2012).

De acordo com o especialista as previsões se mostraram corretas em 96% dos resultados. Com relação à tecnologia usada na campanha ao avaliar o seu uso, Slaby¹⁰⁰ diz que a tecnologia representou uma força de multiplicação para toda a organização.

A tecnologia e a análise de dados tornaram nossos voluntários mais eficientes, pois permitiu que se comunicassem melhor uns com os outros e evitassem perder tempo indo atrás dos alvos errados. Não tínhamos a intenção de construir a tecnologia mais

¹⁰⁰ Em entrevista a Revista IstoÉ do dia 11 de junho de 2013.

avançada, mas usá-la para tornar a campanha mais eficiente (SLABY, entrevista para a Revista Isto É, 2013).

Para dar conta do enorme volume de dados dos eleitores americanos e situações da campanha e seus adversários, a campanha Obama for America (OFA) em 2012 utilizou diversas tecnologias. Entre elas estão:

- (1) Amazon WebServices (AWS) para armazenar e processar o enorme volume de dados;
- (2) Ferramentas de banco de dados como Amazon DynamoDB e Amazon RDS;
- (3) Utilização de arquitetura orientada a serviços (criação da Narwhal).

A Amazon WebServices oferece um conjunto de serviços de computação remota que, juntos constituem uma plataforma de computação em nuvem. A opção pela computação em nuvem na campanha OFA veio da necessidade de armazenamento de dados numa central virtual para evitar duplicação, envolvendo em seus bancos de dados voluntários, doadores e eleitores ao mesmo tempo. A computação em nuvem oferecia essa solução para a escala de produção que a campanha precisava. Com a AWS foi possível construir cerca de 200 aplicações com um conjunto de APIs (aplicativos) que alavancavam diversos conjuntos de dados e servia como uma fonte de dados compartilhada.

Uma das soluções foi o CallTool, um aplicativo que permitiu que milhares de voluntários fizessem milhões de chamadas telefônicas para os eleitores durante os últimos quatro dias de campanha.

Com as restrições de tempo e a necessidade de agilidade a campanha OFA construiu uma série de aplicativos (APis) que expuseram os dados da campanha interna de uma forma que novas aplicações vindas de seus colaboradores pudessem facilmente ser consumidas.

A campanha de Obama também utilizou ferramentas de CRM (Customer Relationship Management) que é um tipo de software que gerencia clientes e potenciais clientes de uma organização. O CRM automatizava diversas tarefas como lembrar de ligar para determinados contatos em períodos de tempos regulares e também mandava automaticamente e regularmente mensagens de para os cadastrados. Ferramentas modernas de CRM estão disponíveis na nuvem e podem integrar o banco de dados do site com as informações das redes sociais e além de gerenciar e organizar toda ação dos colaboradores, também unifica as informações em um ponto central para análise (MARTIN, 2014).

Um dos sucessos da campanha foi a criação de *Narwhal*, um programa que gerenciava e integrava várias fontes de dados. Os dados e informações que eram recolhidos nas comunicações digitais eram associados aos dados coletados por aqueles que trabalhavam em operações de campo como telefonemas, recrutamento voluntário, captação de recursos e ações

porta a porta (ISSENBERG, 2013). O Projeto *Narwhal*, como era chamado por especialistas e analistas de campanhas eleitorais americanas, descreve o esforço da equipe de Obama para conectar os bancos de dados anteriormente separados, de modo que a informação sobre os potenciais eleitores se torna acessível para vários ramos da campanha. Issenberg (2013) chama *Narwhal* de mistura de múltiplas identidades do cidadão – o ativista engajado on-line, o eleitor off-line, o doador, o voluntário – em único perfil político e unificado. O projeto, segundo o jornalista, era tão detalhado que quando um doador atingisse o limite de contribuição de U\$ 2.500,00 ele era persuadido a se voluntariar ao invés de doar mais. Além disso, cada vez que um voluntário individual oferecia para ajudar com fundos para o presidente, ele era solicitado a fazer *login* no site de reeleição com suas credenciais do Facebook¹⁰¹. Assim, seu perfil era direcionado para o Facebook Connect, que é a interface digital que compartilha informações pessoais de um usuário com um terceiro. Conscientemente ou não o eleitor acabava por fornecer informações públicas da sua rede como sexo, idade, data de nascimento, gostos e preferências, bem como suas listas de amigos. Com isso o programa encontrava on-line e off-line padrões de comportamento de informações de eleitores, com a finalidade de personalizar cada interação que o eleitor tinha com a campanha.

5.5.3 A segmentação da comunicação na campanha

A segmentação das bases da campanha OFA permitiu que comunicados específicos fossem enviados por SMS (cerca de 1 milhão de cadastrados), com o objetivo de promover ações de ativação. Como por exemplo, por meio do recurso “porta em porta”, os voluntários batiam na casa de eleitores em seus bairros, após um mapeamento inicial feito pela equipe de campanha. O morador da casa era um idoso, com posse de armas e muito religioso? Então nem adiantava tocar a campainha ou enviar uma mala direta – provavelmente era um republicano empedernido e irrevogável. Já se fosse uma mulher solteira ou viúva, de origem latina, ligada a algum sindicato e ainda não estivesse no banco de cadastro, a campanha tentava conquistar mais um voto. A campanha tinha um mapeamento muito detalhado, dos tipos de eleitores e seus gostos e comportamentos eleitorais. Numa pesquisa¹⁰² feita pela empresa americana Engine Yard, por exemplo, que relaciona o perfil do eleitorado com os

¹⁰¹ O Facebook na campanha de 2008 tinha 40 milhões de usuários nos Estados Unidos. Em 2012 eram 160 milhões que incluía quase todo o público votante americano.

¹⁰² Pesquisa: Eleitores de Obama têm iPhone e usuários de Android preferem Romney. Portal IG. Disponível em: <http://ultimosegundo.ig.com.br/mundo/2012-10-17/pesquisa-eleitores-de-obama-tem-iphone-e-usuarios-de-android-preferem-romney.html>. Acesso em: 4 de Mar. 2014.

dados de mobile, indicou que usuários de iPhone tendiam a votar no Partido Democrata enquanto os usuários de Android tendiam a votar em Mitt Romney. Segundo a pesquisa 88% dos americanos têm celulares e destes 53% possuem smartphones. A pesquisa também mostrava os gostos musicais de cada um desses usuários.

O grau de detalhamento da campanha a partir das análises de dados era expressivo na campanha, que era possível saber até as informações mais pessoais do cidadão americano. Como demonstra o exemplo a seguir relatado por Moraes.

Imagine uma mãe de duas crianças que vive em uma cidadezinha no estado de Ohio, no centro-oeste dos Estados Unidos. Ela votou na última eleição, cadastrou-se e navegou algumas vezes no site de Obama (menos do que na disputa de 2008), mas nunca doou. Seus filhos estudam em escola pública. Ela costuma tuitar sobre o meio ambiente e ainda mantém uma página no Facebook sobre comida orgânica. Tudo isso ficaria registrado no banco de dados da equipe de Obama. “A campanha mandaria para ela e de Michelle Obama sobre as políticas ambientais planejadas pelo presidente e sobre educação pública”, afirma o consultor Andrew Rasiej. Apenas mulheres com perfil semelhante receberiam esse tipo de mensagem. (MORAES, 2013)¹⁰³

A preocupação da mensagem do candidato está sempre veiculada aos interesses de cada eleitor. Assim, segundo Moraes, dificilmente algum americano corria o risco de receber alguma forma de divulgação de Obama, se realmente não fosse do seu interesse.

O uso do *Google Analytics* na campanha contribuiu para a otimização rápida, em tempo real, especificamente durante os debates presidenciais e vice-presidenciais. Uma pesquisa¹⁰⁴ realizada na época pela *Global Strategy Group* apontava que a internet oferecia aos candidatos a oportunidade real para alcançar os eleitores indecisos. E seguindo a pesquisa a equipe colocou em prática suas estratégias de análise de dados e formas de comunicar mensagens para públicos específicos com conteúdos do seu interesse. Para trabalhar a comunicação e mensagens segmentadas, a equipe tinha pesquisas detalhadas sobre os grupos de eleitores e suas preferências pelos meios de comunicação.

Nessas pesquisas com consumidores descobriu-se que se as pessoas que assistem 30 Rock e beber Molson são mais propensas a ser democratas e que votam com frequência. Já aqueles que bebem Coors Light e assistem a NCIS são provavelmente republicanos leis nos quais as estratégias de comunicação de Obama não iriam surtir nenhum efeito. Isso ajuda campanhas a decidir onde anunciar e para quem anunciar (ISSENBERG, 2013). A campanha

¹⁰³ Retirado do artigo: Big data ajudou Obama a ganhar eleição. Disponível em: http://info.abril.com.br/noticias/ti/big-data-ajudou-obama-a-ganhar-eleicoes-15012013-25.shl?utm_source=messiaslg&utm_medium=linkedin&utm_campaign=Feed%3A+NoticiasINFO-TI+%28Not%C3%ADcias+INFO+-+TI%29. Acesso em: 20 Ab. 2015.

¹⁰⁴ De acordo com a pesquisa, 500 eleitores foram entrevistados na Flórida, Ohio, Nevada, Virgínia e Wisconsin. A pesquisa foi realizada entre 02 de outubro e 07 de outubro de 2012.

buscou atingir também os telespectadores de TV a cabo, buscando saber sobre as audiências de canais a cabo pequenas, em que os anúncios eram menos caros. Um dos exemplos citados pela equipe de campanha na época era de que existiam mais maneiras do que as notícias locais para alcançar os eleitores americanos. Como por exemplo os telespectadores de HDNet (que conta com shows de rocks e reprises “Cheers”) e o canal Boomerang, que é um canal de desenhos animados para crianças, tendem a votar apenas em eleições presidenciais. E foi assim que todas as propagandas da campanha de Obama em 2012 foram pensadas e direcionadas.

Ao usar relatórios em tempo real, a equipe da campanha também percebeu rapidamente os termos de busca relacionados ao debate e trouxe a maioria dos espectadores para o site de Obama. A quantidade maciça de dados e informações possibilitou visualizar comportamentos e orientar ações. Todas as informações serviram para adaptar as estratégias da campanha de um modo detalhista, a fim de comunicar mensagens diretamente de acordo com as necessidades de cada grupo de eleitores americanos. Assim, eram colocados anúncios de busca nestes termos no Google para direcionar os usuários a fatos sobre a posição do presidente em questões-chaves. A Figura 13 mostra como eram feitos os anúncios on-line em cada estado considerado decisivo para a campanha.

Figura 13: Anúncio no Orlando Sentinel da Flórida

The image shows a screenshot of the Orlando Sentinel website on Tuesday, Nov. 6, 2012, at 1:22 AM EST. The temperature is 64° F. The website features a navigation bar with links for HOME, LOCAL, WEATHER, SPORTS, BUSINESS, ENTERTAINMENT, LIFE/FAMILY, HEALTH, TRAVEL, OPINION, VIDEOS, and PAY LESS. Below the navigation bar is a trending section with links for Red Cross, Amanda Bynes, James Bond, Watering Restrictions, Sharon Osbourne, Dreamliner Flight, Sam Worthington, and George Zimmerman. The main content area displays a large banner for Election Day in Florida, featuring a photo of Barack Obama and the text: "ELECTION DAY IS TODAY!", "POLLS CLOSE AT 7PM", and "CONFIRM YOUR VOTING LOCATION". Below the banner are four smaller tiles: "VOTER GUIDE 2012", "Election 2012: ENDORSEMENTS", "Election 2012: Understanding", and "FLORIDA VOTES ELECTION DAY".

Fonte: Orlando Sentinel (2012)

Outro exemplo de segmentação e direcionamento de anúncios foi quando o governador Mitt Romney usou termos relacionado às mulheres americanas durante o segundo debate presidencial¹⁰⁵. A equipe de campanha de Obama começou a mostrar anúncios quase que instantaneamente, nas principais mídias sociais e na sua página oficial, com conteúdos que ligavam a um fato descrevendo a opinião do presidente sobre a igualdade das mulheres.

Lubin (2013) diz que devido ao ritmo e à complexidade das campanhas publicitárias que estavam em execução, a equipe de marketing digital precisava usar um site de publicidade consistente e confiável com plataforma centralizada de relatórios para coletar e disseminar dados sobre uma base diária. Com isso, foi adicionando em seus relatórios a capacidade de analisar o conteúdo de blogs, postagens Tumber e outras mídias sociais.

Segundo Issenberg (2015) a campanha de Obama em 2012 foi capaz de observar quem *tweetou* sobre um determinado problema ou feito um post no *Facebook* ou blog sobre a campanha. Essas informações reforçaram os modelos estatísticos, e pode ser combinado aos registros de eleitores individuais, dando a equipe de campo mais informações fundamentadas para trabalhar as estratégias persuasivas.

A segmentação também era vista em outros meios de comunicação como em Ohio, onde a campanha produziu uma revista chamada “Obama”, no estilo da antiga *Life*¹⁰⁶. O alvo eram os eleitores idosos. Na capa da revista estavam os pais de Obama. Ele, veterano do Exército. Assim, a campanha mexia com a nostalgia da Segunda Guerra Mundial, para tocar o coração dos eleitores – mostrando que Obama era alguém próximo deles.

5.5.4 As redes sociais na campanha

Obama em 2012 estava presente em 18 mídias sociais entre elas: Facebook, MySpace, Youtube, Flickr, Digg, Twitter, Eventful, LinkedIn, Block, Planet, Faithbase, Eons, Glu, MiGent, MyBatanga, AsianAve e DNC Party buieder. Cada um desses sites correspondia a um nicho de eleitores como MiGente e Batanga era voltado para os hispânicos; BlackPlanet era para os afro-americanos; Eons para a melhor idade; FaithBase para cristãos, e assim as mensagens para essas redes na campanha eram enviadas de forma segmentada.

Além disso, todas as comunicações sobre a agenda da campanha, as reuniões com voluntários, os discursos do presidente, fotos, vídeos, foram distribuídos por meio das redes sociais. Obama manteve contas em uma variedade de plataformas, incluindo Twitter,

¹⁰⁵ Disponível em: <http://noticias.uol.com.br/internacional/ultimas-noticias/2012/10/16/mais-agressivo-obama-reage-e-ataca-romney-no-2-debate-presidencial-dos-eua.htm>. Acesso em: 02 Mar. 2015.

¹⁰⁶ The George Washington University. <https://efeitoobama.wordpress.com/2009/10/17/o-mapa-do-eleitor-americano-segundo-os-democratas/>. Acesso em: 02 Mar. 2015.

Foursquare, Facebook e Tumblr. O primeiro post no Tumblr da campanha pediu aos eleitores que enviassem seus próprios conteúdos para o blog, uma forma de colaboração e captação de informações e dados para a campanha.

Este era o post:

We'd like this Tumblr to be a huge collaborative storytelling effort -- a place for people across the country to share what's going on in our respective corners of it and how we're getting involved in this campaign to keep making it better¹⁰⁷.

Ao seguir a página oficial do presidente (www.barackobama.com) o internauta automaticamente transferia diversas informações importantes diretamente à equipe de Obama, como sua idade, localização, escolaridade e profissão. Enquanto o adversário republicano Mitt Romney trabalhava com o tradicional método de pesquisa, sabendo em termos gerais a opinião de seus eleitores, Obama tinha dados fundamentais certos e microsegmentados (ROSSI, 2012).

Segundo a análise feita pelo jornal americano *Washington Post* sobre a campanha de Obama em 2012, a atualização do site oficial da campanha OFA foi feita com artigos e notícias de qualidade em ritmo quase frenético. Além disso, as atualizações eram publicadas por estado, ou seja, diversas equipes criavam conteúdos relevantes em toneladas para publicações no site de maneira personalizada. O site foi fundamental na campanha de Obama porque era a estrutura básica de conteúdo disseminada via mídias e redes sociais, no entanto, a estrutura central da estratégia era visivelmente alcançar os objetivos de conversão que eram o cadastro de eleitores e doações para a campanha.

Assim, redes sociais digitais tiveram um papel fundamental na proliferação das informações de campanha no âmbito digital, se transformando no evento político mais tuitado da história dos EUA¹⁰⁸. Exemplo disso foi a foto do presidente abraçado à mulher, Michelle Obama, comemorando a vitória que teve mais de 630 mil retweets, com mais de 20 milhões de mensagens, confirmando a influência das redes sociais na opinião pública. A foto foi postada acompanhada da frase "*Four More Years*" (em português- *Mais Quatro Anos*) (Figura 14).

¹⁰⁷ Tradução livre: "Gostaríamos que este Tumblr seja um grande esforço colaborativo em contar histórias - um lugar para pessoas em todo o país compartilharem o que está acontecendo em nossos respectivos espaços e como nós estamos nos envolvendo nesta campanha para continuar a melhorá-la."

¹⁰⁸ Informações disponível em: <<http://www.portugues.rfi.com/america/20121107-obama-o-tuiteiro-mais-curtido-do-mundo>>. Acesso em: 20 Ag. 2014.

Figura 14: Página Oficial do Twitter de Obama



Fonte: Twitter oficial Obama (2012)

No Facebook, a campanha de Obama teve 3 milhões de curtidas, pulverizando todos os recordes da maior rede social do mundo (STIVANIN, 2012). Segundo a avaliação do The Guardian (2012) a campanha de Obama usou o poder do Facebook para direcionar os eleitores individuais. A rede estava com isso se transformando pela primeira vez numa grande ferramenta de campanha que tinha o potencial de transformar a amizade em uma arma política.

O exemplo da campanha OFA apresentado ilustra de forma prática como campanhas eleitorais podem obter resultados com ações a partir dos usos de ferramentas do *big data* e da metodologia do *microtargeting*, servindo como um grande aliado nas estratégias de comunicação e marketing eleitoral. Essa campanha serve como case para tentarmos entender como campanhas eleitorais brasileiras tentam aproximar suas estratégias a partir do uso do *big data*, bem como de outras ferramentas e metodologias que privilegiem o uso de grandes bases de dados e informações.

PARTE III – A PESQUISA DE CAMPO

6 VISÕES SOBRE OS USOS DO *BIG DATA* EM CAMPANHAS ELEITORAIS BRASILEIRAS

Este capítulo é dedicado à pesquisa empírica, que tem como objeto de investigação os usos do *big data* em campanhas eleitorais no contexto brasileiro. Aqui são apresentados a metodologia que envolveu toda a pesquisa de campo, as técnicas utilizadas na coleta de dados, as etapas da pesquisa e os resultados da pesquisa de campo, com descrição de cada etapa da pesquisa de forma detalhada. Por fim, são apresentadas as considerações da pesquisa de campo em cada fase da investigação.

6.1 METODOLOGIA DA PESQUISA

Para Gil (2008) qualquer classificação de pesquisa deve seguir algum critério e a partir daí será norteado todos os objetivos a serem alcançados numa investigação científica. Dessa forma, a pesquisa empreendida neste estudo é de caráter exploratória e qualitativa (GIL, 2008). A pesquisa exploratória tem como objetivo levar o pesquisador a conhecer mais sobre uma temática que ainda é pouca conhecida tanto no campo acadêmico quanto prático, como é o caso do uso do *big data* em campanhas eleitorais.

Como bem observa Gil (2008), ao final de uma pesquisa exploratória, o pesquisador conhecerá mais sobre aquele assunto, e estará apto a construir hipóteses. A pesquisa exploratória depende da intuição do explorador (neste caso, da intuição do pesquisador). Por ser um tipo de pesquisa muito específica, quase sempre ela assume a forma de um estudo de caso (GIL, 2008). Como qualquer pesquisa, ela depende também de uma pesquisa bibliográfica, pois mesmo que existam poucas referências sobre o assunto pesquisado, nenhuma pesquisa hoje começa totalmente do zero. Haverá sempre alguma obra, ou entrevista com pessoas que tiveram experiências práticas com problemas semelhantes ou análise de exemplos análogos que podem estimular a compreensão. Como será apresentado logo a frente, optou-se neste trabalho dividir a pesquisa em três fases que se complementam com o objetivo geral deste trabalho.

6.1.1 Objetivos da pesquisa

A pesquisa tem por objetivos:

- (a) Conhecer os trabalhos que empresas de dados brasileiras estão fazendo em relação ao uso do *big data* em processos eleitorais;**
- (b) Coletar opiniões e visões de especialistas sobre o uso de *big data* em campanhas eleitorais;**

- (c) **Identificar se já existem experiências no uso do *big data* em campanhas eleitorais no Brasil;**
- (d) **Observar e analisar uma campanha que use o *big data* como estratégias em suas ações de comunicação e marketing eleitoral.**

6.1.2 Instrumentos e técnicas de pesquisa

Para o desenvolvimento da presente pesquisa foram utilizados os seguintes instrumentos:

- **Entrevistas**

Optou-se por trabalhar com dois tipos de entrevistas: a semiestruturada e a não estruturada (MANZINI, 2004). A entrevista semiestruturada é guiada pelo roteiro de questões, o qual permite uma organização flexível e ampliação dos questionamentos à medida que as informações vão sendo fornecidas pelo entrevistado. A entrevista não estruturada é aquela que oferece ampla liberdade na formulação de perguntas e na intervenção da fala do entrevistado.

As entrevistas tiveram o propósito de esclarecer mais sobre a temática e o fenômeno *big data* em campanhas eleitorais. Foram entrevistados especialistas com diferentes visões sobre a temática.

- **Observação de campo**

Segundo Alvarez, a observação é o “único instrumento de pesquisa e coleta de dados que permite informar o que ocorre de verdade, na situação real, de fato” (ALVAREZ, 1991, p.560).

A observação foi usada neste trabalho como critério para verificar a veracidade das informações obtidas a partir de entrevistas realizadas com especialistas sobre os usos do *big data* em campanhas eleitorais, bem como na pesquisa de campo dentro do comitê político da campanha eleitoral de 2014 para governo do Estado do Rio de Janeiro.

A observação direta apresenta algumas vantagens para o pesquisador em pesquisas exploratórias que valem ser destacadas:

- (1) Proporciona a experiência direta com o fenômeno estudado;
- (2) Possibilita recorrer aos conhecimentos e experiências pessoais como auxiliares no processo de investigação;
- (3) Permite chegar mais perto das “perspectivas dos profissionais envolvidos na campanha eleitoral”;
- (4) Possibilita “descobrir” aspectos novos dos usos do *big data* em campanhas eleitorais;

- (5) Permite a coleta de dados em situações em que seriam impossíveis outras formas de comunicação.

6.1.3 Coleta de dados e análise dos resultados

A coleta dos dados se deu a partir das pesquisas bibliográficas e documentais, entrevistas e da observação de campo. Depois de realizadas as entrevistas com os especialistas elas foram transcritas, respeitando todas as falas dos pesquisados. Após a transcrição das entrevistas foi feita a análise de conteúdo proposta por Bardin (1995), que consiste na leitura detalhada de todo o material transcrito, na identificação de palavras e conjunto de palavras que tenham sentido para a pesquisa. Assim como na classificação em categorias ou temas que tenham, semelhanças quanto ao critério sintático ou semântico.

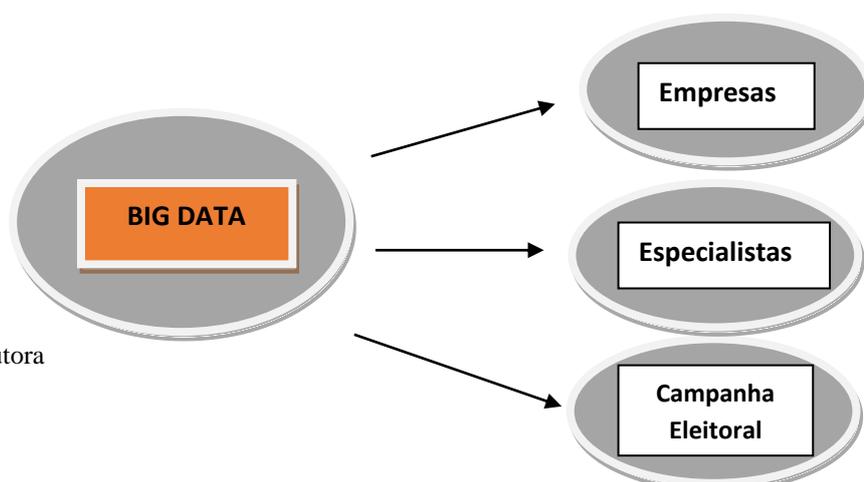
As análises dos dados coletados foram primariamente qualitativas, mas, para a compreensão dos resultados, os dados foram quantificados e representados por meio de quadros e tabelas. Portanto, o estudo abrange análises quantitativas, ainda de forma secundária às análises qualitativas.

Por ser uma pesquisa qualitativa não se empregou o uso de instrumentos e técnicas estatísticas sofisticadas no tratamento dos dados e sim análise de conteúdo e inferências a partir das falas dos entrevistados e das observações da pesquisa de campo sobre o objeto investigado.

6.1.4 Fases da pesquisa de campo

Para alcançar os objetivos, a pesquisa de campo se deu em três fases: Fase 1 – Pesquisa em empresas de dados brasileiras; Fase 2- Entrevista com especialistas; Fase 3- Observação direta em uma campanha eleitoral.

Figura 15: Fases da pesquisa de campo



Fonte: Autora

Fase 1 – Pesquisa em empresas de pesquisa e dados brasileiras

Nesta fase o objetivo da pesquisa foi entender como as empresas de dados estão trabalhando com *big data* no campo da política eleitoral. Buscou-se entender como aproveitam os dados e informações gerados, principalmente nas mídias sociais; que tipos de serviços oferecem para as campanhas eleitorais; quais metodologias, ferramentas e tecnologias usam; quais especialidades de profissionais são necessárias para trabalhar com o *big data*.

As pesquisas foram realizadas em duas empresas brasileiras, a ELife e a Ideia Inteligência, no período compreendido entre dezembro de 2013 a julho de 2014, nas cidades de São Paulo e Rio de Janeiro.

A amostra selecionada não é probabilística e intencional e a escolha recaiu no fato de que tais empresas trabalham na área de análises de dados voltados para campanhas eleitorais e/ou processos eleitorais, além da facilidade de acesso e contatos para a pesquisa de campo. Durante essas visitas foram realizadas entrevistas semiestruturadas com profissionais responsáveis pelas atividades da empresa. Ao todo foram realizadas 5 entrevistas, sendo 2 na empresa ELife (Sócio proprietário da empresa e com responsável pelos desenvolvimento de projetos) e 3 na empresa Ideia Inteligência (Gerente de Marketing, Cientista Política e profissional de Tecnologia da Informação). As entrevistas foram gravadas com o consentimento dos entrevistados e foi preparado um caderno de anotações com o roteiro semiestruturado.

Ao longo da análise dos dados coletados nas empresas, percebeu-se que à medida que se caminhava pelo entendimento das suas atividades novos questionamentos afluíam que precisavam ser esclarecidos. Dessa forma, para sanar as dúvidas manteve-se contatos com as empresas pesquisadas e os profissionais entrevistados via telefone e correio eletrônico (e-mail).

As ilustrações que correspondem às empresas no corpo do trabalho foram disponibilizadas pela equipe de marketing da empresa para a pesquisa, a fim de dar uma visão mais clara dos tipos de serviços e metodologias usadas em suas atividades.

Fase 2- Entrevista com especialistas

Foram realizadas entrevistas com profissionais, especialistas e pesquisadores na temática *big data* e campanhas eleitorais. O critério para a escolha dos entrevistados seguiu o caráter de representatividade, seja no meio acadêmico ou campo profissional relacionado à temática

da pesquisa, e também pela oportunidade e disponibilidade dos entrevistados em contribuir com o trabalho. A seleção começou por profissionais ligados ao tema no contexto brasileiro e depois partiu-se para uma investigação internacional. Para identificar esses profissionais foi feito um mapeamento prévio na internet e em ambientes acadêmicos, por meio de suas publicações científicas e em blogs da área da temática. A partir daí foram estabelecidos os primeiros contatos via correio eletrônico (e-mail) e por telefone para os agendamentos das entrevistas.

As entrevistas se deram de três maneiras: pessoalmente (face a face), via e-mail e Skype. Tanto pessoalmente quanto por Skype, as entrevistas tiveram em média a duração de 30 minutos a uma hora e foram gravadas, posteriormente transcritas e arquivadas, para que fossem consultadas sempre que houvesse necessidade no decorrer da escrita da tese.

As entrevistas partiram de um roteiro semiestruturado entre 8 a 12 questões abertas, dependendo da especialidade do profissional e foram realizadas de dezembro de 2013 a abril de 2015. Todos os entrevistados foram identificados mediante autorização. Foram entrevistados um total de 12 especialistas, sendo 10 brasileiros e 2 americanos. Essas entrevistas totalizaram 350 minutos de depoimentos gravados (Skype e pessoalmente). Na análise do conteúdo das entrevistas, a escolha se deu pelos trechos potencialmente relevantes para a temática da pesquisa. Em seguida, o conteúdo foi analisado em três níveis:

- 1) **Nível descritivo:** no qual os posicionamentos dos entrevistados foram descritos;
- 2) **Nível correlacional:** no qual essas posições foram associadas e comparadas entre os entrevistados;
- 3) **Nível interpretativo:** no qual se procurou descobrir as relações nas reações dos entrevistados a respeito da temática.

Não foi empregado nenhum tipo de tecnologia ou *software* específico para fazer a análise das entrevistas, foi tudo feito manualmente e seguindo a linha de classificação e interpretação das falas. Para um melhor entendimento e relação dos depoimentos dos especialistas, optou-se por dividir e organizar as entrevistas em tópicos relacionados à temática da investigação que foram divididos em: **Mudanças com o *big data*; O *big data* em campanhas eleitorais; Estrutura e especialidades profissionais; Usos de dados pessoais; Pontos negativos do *big data*.**

Fase 3 – Observação direta em uma campanha eleitoral brasileira

Foi realizada uma pesquisa de campo em um comitê de campanha eleitoral no processo eleitoral de 2014, na disputa para o governo do Estado do Rio de Janeiro. O objetivo

dessa pesquisa era conhecer de perto como se dava uma campanha eleitoral com o uso do *big data* e verificar em que medida o *big data* era usado nas estratégias de informação e comunicação do candidato. A escolha da campanha se deu pela oportunidade da pesquisadora ser inserida na equipe de comunicação e marketing do candidato.

Para realizar essa pesquisa, foi escolhido o método de observação direta não estruturada. A técnica possibilitou coletar dados sobre a campanha eleitoral e observar a dinâmica de trabalho das equipes de informação e das equipes de comunicação na campanha.

As entrevistas não estruturadas se deram à medida que acompanhava as atividades dos profissionais. Não seguiu uma rigidez de perguntas e nem foram gravadas, foi usado apenas caderno de anotações para registrar os principais pontos dos depoimentos dos entrevistados, na medida que os questionamentos eram respondidos. Optou-se por esse tipo de entrevista para que não fosse interrompida as atividades que estavam sendo desenvolvidas pelos profissionais, já que numa campanha o trabalho é intenso e o tempo é limitado.

6.2 EMPRESAS BRASILEIRAS DE PESQUISA E DADOS

6.2.1 Elife

A escolha pela empresa Elife para ser pesquisada foi direcionada a partir das inúmeras publicações encontradas na internet sobre a suas atividades no uso de análise de *big data* e aplicação de metodologias para atender o mercado de consumidores e também o político eleitoral.

A empresa Elife está localizada em São Paulo (SP) e foi criada pelo jornalista Alessandro Barbosa Lima em 2004, a partir do monitoramento da rede Orkut. Alessandro Lima é um dos sócios diretores da empresa Elife e atua diretamente no negócio de mídias digitais com diversos tipos de serviços oferecidos principalmente para o mercado empresarial. A companhia iniciou suas atividades em 2006 e conta com duas divisões: a Elife Monitor, responsável pelo monitoramento, análise e *insights* em redes sociais utilizando tecnologias próprias, e a Elife Social CRM, para gestão do relacionamento em redes sociais. O grupo Elife abrange também outras duas empresas: a Social Agency (agência de marketing on-line focada no posicionamento da marca em redes sociais), com operação totalmente independente, e a Landscape (especializada em ações que utilizam projeção e/ou realidade aumentada).

A ideia da empresa surgiu a partir da pesquisa acadêmica de Lima no curso de mestrado na Escola de Comunicações e Artes da USP sobre comunicação interpessoal on-line.

Segundo Lima (2014), a partir daí foi possível ter uma leitura mais aprofundada sobre o fenômeno das redes sociais e sobre a topologia destas redes.

“Dois fenômenos se mostraram para mim, naquele momento, irreversíveis. O crescimento exponencial das redes sociais – em comunidades que representam parte das nossas redes sociais, como o Orkut entre outras” (LIMA, 2014).

As oportunidades de mercado nessa área são destacadas pelo diretor da empresa a partir da sua observação, de que a comunicação boca a boca - um fenômeno até então invisível e não mensurável – passava com a existência das redes sociais, tornando-se visível e mensurável.

“A internet torna o boca a boca, seja ele negativo ou positivo um fenômeno transparente. Com isso ganham primeiros os consumidores – que podem buscar a opinião sobre qualquer produto ou serviço e ganham as empresas que podem compreender melhor as necessidades dos consumidores” (LIMA, 2014).

Lima explica que a proposta da empresa é analisar o “discurso coletivo”. Não existe interesse em saber individualmente a opinião de nenhum consumidor ou entender em profundidade um indivíduo ou como ele navega. Segundo Lima (2014) isso seria uma invasão de privacidade dos indivíduos.

“Por isso nosso posicionamento é a monitoração e análise da comunicação boca-a-boca da multidão, que chamamos de e-life. Esta multidão escreve posts, comentários, publica vídeos e podcasts on-line a cada segundo. Esta multidão busca informações sobre produtos e serviços, mas também fornece informações sobre experiências com produtos que comprou” (LIMA, 2014).

O monitoramento das mídias sociais é feita de forma automatizada, usando *software* específicos, mas depende também das análises qualitativas dos especialistas.

“O trabalho é diário, 24 x 7. As máquinas varrem de madrugada a web e os analistas avaliam os resultados durante o dia. Temos de tudo. Desde de análises aprofundadas de uma indústria, marcas e concorrentes até prevenção de crises em processo embrionário” (LIMA, 2014).

No que se refere às atividades ligadas a campanhas ou processos eleitorais, o coordenador de projetos da empresa Breno Sotto¹⁰⁹ explica que a Elife faz armazenamento de informações da internet e com elas classifica diferentes hábitos e padrões nas redes sociais.

“Nossa intenção com esse tipo de trabalho é ajudar os profissionais de comunicação e marketing na elaboração de suas estratégias de campanha política” (SOTTO, 2014).

Sobre isso Sotto ainda completa:

¹⁰⁹ Breno Sotto deu entrevistas por telefone em novembro de 2014 para a autora desta tese.

“O monitoramento das redes sociais sempre foi muito focado na marca, mas estamos virando para outra direção para monitorar o que se fala do consumidor/eleitor. Monitorando grupos específicos, podemos observar de que marcas ou candidatos essas pessoas falam. Vai ajudar as empresas de marketing político a ter uma visão mais aprofundada de quem é o eleitor” (SOTTO, 2014).

No que se refere a isso, a empresa Elife estabelece suas pesquisas em duas etapas:

- 1) coleta de dados em diversos ambientes da web, a partir de *softwares* próprios, que coletam e organizam os dados em relatórios específicos;
- 2) análise de dados, feitas por analistas de mercado, no intuito de interpretar os dados coletados.

Faz parte dessa segunda etapa a codificação das manifestações on-line dos consumidores como: positivos, negativos, neutros e mixed (que misturam conotações positiva e negativas). Isso é chamado de mineração de textos.

Sotto explica como funciona o monitoramento.

“O Buzzmonitor permite ver e acompanhar as timelines de usuários do Twitter e, já no passo seguinte, organiza a informação obtida através da criação de reports específicos e completamente customizáveis, o que facilita a visualização de dados e nos ajuda a chegar a conclusões importantes para qualquer tipo de mercado, seja político ou comercial” (SOTTO 2014).

Ainda sobre a ferramenta de análise, Sotto diz que automaticamente, o Buzzmonitor classifica para apurar características do público-alvo através dos universos semânticos das conversas e perfis dos usuários.

“Criamos relatórios temáticos no Buzzmonitor para separarmos as informações relevantes sobre a nossa base de dados. Aqui a análise semântica, já programada na plataforma, é uma grande ajuda, pois associa universos significativos e agrupa termos com características para nos dar resultados de forma organizada e conclusiva” (SOTTO, 2014).

A Elife atende diversas tipos de empresas e marcas, e em novembro de 2013 lançou em Brasília os serviços de *big data* eleitoral. Baseado na ferramenta Buzzmonitor¹¹⁰, o serviço *big data* eleitoral coleta, classifica e analisa os desejos da sociedade, o comportamento dos eleitores e dos políticos. O Buzzmonitor é uma plataforma 3 em 1 de redes sociais, que compreende monitoramento, social CRM e social *analytics*.

O serviço é oferecido de duas formas: uma de forma gratuita para até seis mil dados na versão *freemium*, mas também direcionados para candidatos, partidos e governos, para que as equipes de comunicação e marketing possam trabalhar com a análises de grandes volumes de dados e retirar deles *insights* para a campanha e governo. Na campanha eleitoral de 2014, a

¹¹⁰ O BuzzMonitor traz dados quantitativos como: aonde se fala mais sobre a marca, blogueiros com maior relevância (métrica mensurada através dos incoming links).

ELife serviu de apoio principalmente para as pesquisas políticas que eram realizadas pela imprensa brasileira.

“Nas últimas eleições nossos serviços não foram direcionados para os clientes políticos, realizamos um trabalho paralelo que serviu de apoio aos meios de comunicação sobre o processo eleitoral de 2014” (LIMA, 2014).

Dessa forma, o serviço prestado pela Elife incorpora a infraestrutura de *big data* e as ferramentas de *business intelligence* e que são aplicadas às análises de pesquisas quantitativas e qualitativas tradicionais para traçar estratégias políticas.

Segundo Lima (2014) a proposta do serviço *big data* eleitoral é coletar, classificar e analisar os desejos da sociedade, o comportamento de eleitores e políticos e os movimentos sociais que tenham quaisquer relações com o tema. Os resultados dessas análises são direcionadas e comercializadas, tanto para políticos, agências de comunicação e imprensa.

A metodologia utilizada pelo *big data* eleitoral, além de envolver indicadores como pesquisas eleitorais, também faz análise de dados agregados das redes sociais. Faz parte, também, do *big data* eleitoral o monitoramento do Twitter de alguns políticos brasileiros com o objetivo de analisar suas principais citações, participações e interações nessa rede social. O trabalho de monitoramento para o processo eleitoral de 2014, foi feito desde 11 de setembro de 2013 até as eleições de 2014.

A empresa conta hoje com 250 funcionários em áreas de comunicação e marketing, distribuídos em dois escritórios, um em São Paulo e outro no Recife, além de manter profissionais em mais cinco países como Portugal, Austrália, México, Espanha e Angola.

6.2.2 Ideia Inteligência

A busca pela empresa Ideia Inteligência foi incentivada a partir da entrevista realizada com o publicitário e especialista em campanhas eleitorais Renato Pereira em 2014, que é sócio-proprietário da Agência de comunicação e marketing Prole, que fica no Rio de Janeiro. O diretor da empresa Ideia Inteligência Maurício Moura trabalha em parceria com a Prole em muitas campanhas eleitorais brasileiras e internacionais, gerenciando e analisando dados eleitorais. Antes de criar o Ideia Inteligência, Moura prestava consultoria diretamente para Renato Pereira em diversos projetos e para diferentes clientes.

A empresa tem ramificações também em São Paulo, Brasília e Portugal e um escritório em Washington nos Estados Unidos. Nasceu em 2013 com o intuito de prover inteligência aos gestores públicos, instituições de pesquisa e ensino e estratégias eleitorais. Trabalha com o conceito de *big data* desde a coleta até a análise dos dados. Os princípios éticos e credibilidade de informações são garantidos por meio de práticas que regulam a atividade de

pesquisa no Brasil e no mundo. Tatiana Seixas responsável pelo marketing da empresa ressalta a preocupação da Ideia com as regulamentações.

“As pesquisas do Ideia atendem ao código de autorregulação e de ética da Associação Mundial de Profissionais de Pesquisa (Esomar) e também ao código de ética da Associação Brasileira de Empresas de Pesquisa (Abep)” (SEIXAS, 2014).

A empresa atende diversos tipos de clientes como governos, candidatos, além de empresas e instituições de ensino (Figura 16).

Figura 16: Tipos de pesquisas



Fonte: Ideia Inteligência (2014)

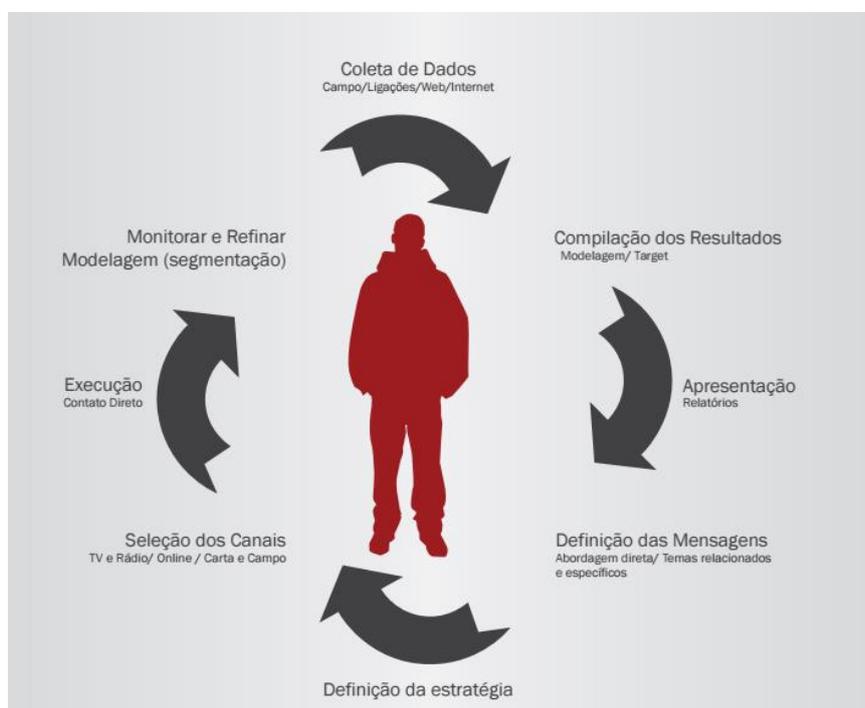
Os partidos políticos e candidatos utilizam as pesquisas e análises feitas pela Ideia Inteligência para avaliar a popularidade e tendências eleitorais, Isso é usado especificamente para testar, mensurar e consolidar suas estratégias de campanha. Entre as atividades estão: análise de imagem de lideranças políticas e sociais; avaliação, gestão e administração pública; pesquisas de opinião pública e de intenção de voto; realiza *tracking* de satisfação (política, social e econômica); oferece subsídios para estratégias/conteúdos de campanha; faz levantamento de imagem e posicionamento de candidaturas.

De acordo com Seixas, a empresa tem como objetivos mobilizar parceiros, analisar dados, fazer diagnósticos sobre informações e opiniões; formular, testar e aprimorar inteligência e estratégias.

“Conduzimos avaliação, diagnósticos e análises que se embasam em informações coletadas através de extenso trabalho de campo, em trackings e no conhecimento profundo de públicos e segmentos alvo. Contribuímos à formulação e aprimoramento de propostas e iniciativas estratégicas. Calculamos o impacto de políticas públicas, iniciativas e estratégias aplicando técnicas de econometria e estatística, como, por exemplo, regressão logística e probabilística e DIFF-in-DIFF” (SEIXAS, 2014).

A Figura 17 demonstra como é o trabalho realizado pelo Ideia Inteligência.

Figura 17: Como trabalha a empresa



Fonte: Ideia Inteligência (2014)

A cientista política Carolina Almeida de Paula trabalha na empresa como coordenadora de pesquisa e conta como o Ideia trabalha com o processo de coleta de dados.

“Não trabalhamos com dados comprados. Os dados são públicos. Parte dos dados eleitorais a gente pega com o TSE e nossos analistas de pesquisas dividem por zonas e seções eleitorais. Agora os dados da população vêm de bancos de dados públicos. A gente não trabalha com dados comprados e privados. Temos bancos variados à medida que vamos trabalhando para cada região. Então o banco varia de Estado para Estado. O banco de dados do Estado do Rio de Janeiro é o banco de dados mais completo que a gente tem, porque fazemos acompanhamento de campanhas desde lá de traz (desde 2008) e atualizamos os bancos constantemente até mesmo fora da época de disputas eleitorais” (PAULA,2014).

A empresa, além de trabalhar com a coleta de dados, por meio de pesquisa, também é responsável, em uma campanha eleitoral, por exemplo, pelas análises de dados, usando técnicas de *microtargeting*, que diz com quem, o que e como a campanha deve falar.

“O princípio fundamental do microtargeting é conhecer profundamente as características e as principais variáveis que influenciam os diferentes grupos de indivíduos, com o objetivo de estabelecer uma estratégia de comunicação personalizada e que impacte cada um dos grupos de forma eficiente. Depois da obtenção das coordenadas geográficas de uma determinada região, é feita uma coleta de dados públicos. Como resultado do estudo, é possível analisar as chances de cada subgrupo votar em um determinado candidato ou comprar um determinado produto, gerando mapas temáticos de calor, disponibilizados para navegação interativa” (SEIXAS, 2014).

Segundo a gerente de marketing, o uso dessa metodologia se configura num diferencial do trabalho realizado com o *big data*, porque busca-se no grande volume de dados

e informações direcionamento certo das ações de campanha política. No setor político, a empresa foi pioneira na aplicação de técnicas de *microtargeting* eleitoral no Brasil. Das 23 eleições brasileiras gerenciadas em 2008, a empresa teve sucesso em 87% delas e de lá para cá já trabalhou em outras campanhas usando essas técnicas e também as análises do *big data*.

Na prática, a empresa faz coleta em cima de dados estatísticos (dados passados de institutos de pesquisa como TSE – Tribunal Superior Eleitoral, TRE – Tribunal Regional Eleitoral e IBGE – Instituto Brasileiro de Geografia e Estatística, PNAD – Pesquisa Nacional por Amostra de Domicílios, entre outros), que mostram o histórico eleitoral e o comportamento do voto e cidadão/eleitor, bem como classifica os eleitores por região, classe social, idade, sexo, renda salarial. Esses dados são armazenados em um único banco de dados. Além disso, trabalham com dados tratados em tempo real que é chamado de *stream computing*¹¹¹, que são dados que vem de coleta e análises de mídias sociais. Essas análises são quantitativas e qualitativas a partir de análise de sentimentos de *Twitter*, *Facebook*. Hoje esses dados estão sendo analisados por uma empresa parceira chamada Casa Digital, que faz parte do grupo da Agência de Publicidade Prole. Lá os profissionais usam *softwares* específicos para fazer monitoramento de opiniões públicas sobre candidatos, políticos, governos e outras instituições.

Para a armazenagem e análise de dados, a empresa usa um *software* que foi desenvolvido pela equipe de tecnologia de informação (TI) chamado de Alexandria, que se constitui num depósito próprio. Segundo o especialista em Tecnologia da Informação Pedro Henrique Reis o *software* é capaz de integrar grandes volumes de dados de diversas fontes e formatos e gerar relatórios em tempo real, de acordo com as necessidades de cada cliente (candidato).

“Nosso principal diferencial é a capacidade de integrar, em tempo real, de maneira única e inovadora, os serviços de coleta de dados (através de diferentes metodologias de pesquisa), análise e geração de inteligência competitiva, com a oferta de uma base de dados estruturada em um sistema de big data. Esta integração de serviços permite que nossos clientes consigam conhecer melhor as características, interesses e necessidades de seu público-alvo, podendo criar estratégias que alcancem melhores resultados, com menor esforço e custo” (REIS, 2014)

O sistema *big data* tem a capacidade de integrar diferentes bancos de dados gerando relatórios e provendo serviços de automação para coleta de dados em diferentes canais, como redes sociais, WhatsApp, e-mail e telefone.

¹¹¹ É um termo que integra continuamente e analisa dados em movimento para a entrega de análises em tempo real. (Informações retiradas do Site da IBM)

A equipe da empresa é multidisciplinar e conta com especialistas de diferentes áreas, como estatística, economia, ciência política, ciências contábeis, direito, tecnologia da informação, publicidade, comunicação, design e jornalismo. No escritório do Rio de Janeiro em períodos eleitorais a empresa chega a contar com 10 profissionais especializados, que ficam no gerenciamento da coleta e análise das informações. Além disso, são contratadas pessoas para o trabalho de pesquisa, que são feitas nas casas e nas ruas com os eleitores.

6.3 CONSIDERAÇÕES DA PESQUISA DE CAMPO

A pesquisa realizada com as duas empresas que trabalham com atividades relacionadas à coleta, armazenamento e análise de dados, a partir de diferentes pesquisas teve por objetivo conhecer como o *big data* vem sendo utilizado no Brasil no contexto eleitoral.

A partir desse trabalho foi possível comprovar a hipótese de que o grande volume de dados e informações que vem sendo gerados hoje pelas pessoas, estão cada vez mais se tornando alvo de experimentações e negociações. Como bem afirma Schonberger-Mayer e Cokier (2013, p. 89), “a matéria-prima do *big data* é a informação”. Informações e dados que sempre foram alvos de coleta e análises passaram a ser vistos por empresas de pesquisas de mercado como um bem primário. Empresas detentoras de *big data* contam com especialistas para extrair valor dos dados e fazem disso um negócio vantagoso. Como se percebeu na pesquisa, as duas empresas têm pouco tempo de mercado no Brasil e já expandiu suas atividades para outras localidades, inclusive fora do Brasil.

Durante pesquisa, foi possível perceber que apesar das empresas trabalharem com o mesmo propósito de atender ao mercado voltado para a política eleitoral, suas atividades e seus serviços são oferecidos de forma diferenciada. Isso comprova que as empresas também estão experimentando novas formas de negócio, como é o caso da Elife que tem seu foco voltado para o mercado consumidor, mas em períodos eleitorais expande suas atividades também para atender o segmento político eleitoral.

A partir da pesquisa realizada é possível montar um quadro comparativo (Quadro 12), estabelecendo diferenças nos serviços oferecidos e atividades desenvolvidas .

Quadro 12: Atividades das empresas de dados

Atividades/Estrutura	Elife	Ideia Inteligência
Dados e Informações	Coleta e análise a partir de redes sociais (Facebook, Twitter)	Coleta a partir de pesquisas de campo (quantitativas, qualitativas) feitas pessoalmente, por telefone, Internet e a partir de mídias sociais.
Serviços	Focada em Inteligência de Mercado; Monitoramento de mídias sociais; Análise de dados quantitativos e qualitativos; Posicionamento da marca/candidato nas mídias sociais; Estratégias de marketing para fortalecimento da marca/candidato.	Focada em coleta de dados e informações por meio de pesquisas de diversas fontes e formatos; Inteligência de Mercado; Consultorias e Projetos Especiais; Soluções de tecnologia.
Metodologia	<i>Brand-centered Monitoring e Consumer-centered Monitoring</i>	<i>Microtargeting</i>
Tecnologia	Buzzmonitor – plataforma 3 em 1 de redes sociais que compreende monitoramento, social CRM, e social Analytics	Sistema Alexandria (Software desenvolvido pela empresa)
Especialidades de profissionais	Analistas de mídias sociais e gerente de projetos	Cientista Político, Cientista social, Analistas de Mídias Sociais, Jornalista, Publicitário, Especialista em Tecnologia da Informação,
Total de Funcionários	250 funcionários distribuídos no Brasil e no exterior	20 funcionários permanentes. No período de campanhas eleitorais o número se altera conforme a necessidade de cada campanha contratante.
Atendimento	Empresas Marcas Produtos Políticos Candidatos	Instituições de ensino públicas Gestores públicos (governo) Agências governamentais Partidos políticos Doadores internacionais (governos) Estrategistas eleitorais e políticos Instituições de ensino privadas Fundações privadas Empresas Doadores internacionais (não-governamentais) Instituições/empresas de pesquisa privadas Instituições de pesquisa públicas Associações e fundações públicas Organizações multilaterais e bilaterais Sindicatos, cooperativas e associações
Extensão	No Brasil: São Paulo e Recife No exterior: Portugal, Espanha, México, Austrália e Angola.	No Brasil: Rio de Janeiro, São Paulo, Brasília. No exterior: Argentina, Venezuela, Portugal, Romênia e Estados Unidos

Elaboração: Autora

Pelo quadro é possível verificar as diferenças entre as duas empresas pesquisadas. A Elife oferece coleta e análise de dados digitais vindos da internet, mídias sociais (Facebook e Twitter) e trabalha apenas nesse ambiente das pesquisas. Já a Ideia Inteligência é uma empresa que trabalha os dados de forma integrada e personalizada desde a coleta dos dados até as suas análises com dados vindos de diferentes fontes e formatos. Suas pesquisas são feitas em ambiente on-line e principalmente off-line.

Assim é possível entender as atividades das duas empresas a partir do ciclo de vida do *big data* proposto por Lastori (2013). A Elife trabalha com a fase de **aquisição** por meio de coleta de dados, porém limita sua coleta em apenas ambientes da internet, trabalhando apenas com dados não estruturados vindos principalmente de redes sociais como o Facebook e o Twitter. Enquanto a Ideia Inteligência faz a coleta em diversas fontes tanto on-line quanto off-line, aqui são coletados dados em redes sociais, blogs, sites, e também de fontes que fornecem dados estruturados como listas públicas, Institutos de Pesquisa (IBGE, PNDE, TRE, TSE) e diretamente através de pesquisas via telefone, pesquisas na rua ou mesmo porta a porta.

No que se refere a segunda fase do ciclo, a **organização**, a Elife organiza os dados coletados em plataformas de armazenamentos. Já a Ideia Inteligência, além de capturar os dados, se preocupa com a limpeza feita por um profissional da estatística, que tem como função classificar e organizar antes de armazenar em bancos de dados no Sistema Alexandria.

Na fase de **análise**, também conhecida como *analytics*, são aplicados os *algoritmos* matemáticos sobre os dados já classificados e armazenados. *Analytics* é definido como inteligência analítica, é a palavra-chave que encerra o ciclo de vida do *big data* de forma tecnológica (NAFFAH, 2012). Através do uso de *analytics* as organizações conseguem acessar, gerenciar e cruzar dados à procura de *insights* de informações contidas neles (VILLELA, 2013). Com *big data*, a diferença é que essa capacidade de se “entender melhor uma situação” se tornou exponencial, uma vez que é possível realizar análises sobre dados volumosos e caóticos, tais como os encontrados na internet, principalmente em ambientes de redes sociais.

Nesse tipo de análise as empresas se utilizam dos conceitos de rastreamento ou monitoramento de dados (BRUNO, 2013; LEMOS, 2012), além disso os *softwares* usados para monitoramento, fazem a mineração de dados (Social CRM) e as análises de sentimentos, principalmente do Twitter e do Facebook (Social Analytics).

Dessa forma, ambas as empresas se utilizam dessa fase em seus negócios, cada uma utiliza a sua plataforma e seu sistema, escolhendo a melhor tecnologia e metodologia para as suas atividades, dando resultados estatísticos de acordo com o perfil de cada cliente ou

campanha eleitoral. Para Villela (2013) através do uso de *analytics* as organizações conseguem acessar, gerenciar e cruzar dados à procura de *insights* de informações contidas neles. Da mesma forma com o uso de *analytics* pesquisadores conseguem analisar informações de milhares de pessoas e compreender o comportamento humano a um nível de população, a partir da introdução de algoritmos¹¹². Assim, as empresas pesquisadas também contam com essas opções em suas análises de *big data*. Essa fase para muitos autores é também chamada de *big data Analytics*.

Na última fase que corresponde a **decisão** as empresas pesquisadas trabalham de forma diferenciada. Essa fase depende dos resultados vindos das análises realizadas. A ELife oferece esse tipo de serviço a partir da agência digital pertencente ao Elife Group. Trabalha de forma independente e é responsável por estratégias de conteúdo, mídia, produção e gestão de presença social de marcas, produtos e serviços, atendendo também candidatos e campanhas. Já a Ideia Inteligência oferece esse tipo de serviço de acordo com o perfil de cada cliente (candidato ou campanha), ou seja, de forma customizada. Se a campanha precisa de todas as etapas desde a coleta até a análise, ou se precisa somente de uma das etapas do processo. Além disso, a empresa Ideia Inteligência também pode oferecer apenas a tecnologia (Sistema Alexandria) ao cliente ou campanha, para que a partir dos bancos de dados do próprio cliente, sejam feitas as análises de *big data*.

Outra diferença que a pesquisa encontrou foi em relação ao foco das atividades e clientes. A ELife é uma empresa voltada para clientes do setor privado e comercial, com extensões para a área política e eleitoral, mas somente em períodos eleitorais. Já a Ideia Inteligência é uma empresa que nasceu voltada para atender principalmente os setores públicos, governos, partidos, candidatos (campanhas eleitorais). Assim, seu foco principal é no segmento político. Durante o período de campanha eleitoral, intensifica suas atividades para essa área e fora do período eleitoral atende governos e setores públicos.

Em relação ao corpo de profissionais, a ELife tem 250 funcionários, distribuídos em seus escritórios pelo Brasil e fora do país. A empresa Ideia Inteligência tem um corpo permanente de 20 profissionais e em épocas de campanhas eleitorais, porém esse número se altera de acordo com as necessidades operacionais de cada cliente (candidato).

¹¹² O *algoritmo* é conceituado como uma sequência finita de instruções bem definidas com a finalidade de resolver um determinado problema computacional. As informações se inserem nesse contexto como os dados de entrada, o *algoritmo* executa o processamento dessas informações e o resultado são as informações. Segundo o dicionário Aurélio um *algoritmo* é um: “processo de cálculo, ou de resolução de um grupo de problemas semelhantes, em que se estipulam, com generalidade e sem restrições, regras formais para a obtenção de resultado ou solução de problema”.

Essa pesquisa nos leva a atender que existe um mercado em expansão no Brasil para cuidar de dados eleitorais. Cada empresa pesquisada demonstrou que esse negócio é viável e está longe de ser dispensável num mercado, que cada vez mais conta com excesso de dados e informações, que se encontram disponíveis para serem coletadas e analisadas, seja para o mundo dos negócios ou para as campanhas eleitorais.

6.4 VISÕES DOS ESPECIALISTAS SOBRE *BIG DATA*

Essa parte da pesquisa de campo apresenta as visões e expectativas que os especialistas têm sobre o uso do *big data*. O Quadro 13 apresenta a lista de profissionais e especialistas que foram entrevistados para esta pesquisa.

Quadro 13: Especialistas entrevistados na pesquisa

ENTREVISTADO	QUALIFICAÇÃO	DATA	FORMA	TEMPO
1. Donald Brown	Diretor do Instituto Big Data da Universidade de Virgínia dos Estados Unidos. Professor de Engenharia e Ciências Aplicadas.	16/12/2013	Entrevista feita pessoalmente	40 min
2. Renato Pereira	Sócio proprietário da Agência de Comunicação e Marketing Prole Antropólogo e Publicitário Gerência de inúmeras campanhas eleitorais brasileiras e internacionais com o <i>microtargeting</i> e <i>big data</i>	18/06/2014	Entrevista realizada pessoalmente	30 min
3. Carolina Almeida de Paula	Dra em Ciência Política pelo Instituto de Estudos Sociais e Políticos da UERJ (2014). Foi coordenadora de pesquisas no IDEIA INTELIGÊNCIA com foco em <i>big data</i> .	24/07/2014	Entrevista realizada pessoalmente	40 min
4. Walter Teixeira Lima Júnior	Jornalista, Professor e Pesquisador na área de <i>big data</i> , redes sociais digitais, campanhas políticas presidenciais na Universidade Metodista de São Paulo. Pós Doutor (USP, Metodista), Doutor em Ciência da Comunicação (USP).		Entrevista realizada por Skype	1h00
5. Carlos Manhanelli	Presidente da Associação Brasileira dos Consultores Políticos. Consultor em Marketing Eleitoral. Jornalista, administrador de empresas, especialista em marketing político. Autor de 11 livros sobre marketing político e estratégias eleitorais.	30/07/2014	Entrevista feita por Skype	40 min

6.Cézar Traurion	Economista, mestre em Ciência da Computação, MBA em Marketing e Serviços. Autor de livros na área de tecnologia da informação, temas como: Big data, software livre, Grid Computing, Software Embarcado, Cloud Computing. Escreve para publicações especializadas, como CIO Magazine, Mundo Java, IMasters. Foi	04/08/2014	Skype	1h00
7.Maurício José Serpa Barros de Moura	Economista (USP), Mestre pela University of Chicago e Doutor em Economia e Política do Setor Público pela FGV-SP, com extensão na George Washington University. Atualmente, é pesquisador visitante da George Washington University em Washington, DC. É um dos diretores da empresa Ideia Inteligência. Experiências com campanhas eleitorais usando <i>microtargeting</i> e <i>big data</i> desde 2008 no Brasil e em âmbito internacional.	23/02/2015	Entrevista realizada por Skype	40 min
8. Christian Conceição Guerreiro Santos	Graduado em Ciência da Computação pela Universidade Federal da Bahia (1998). Atualmente é Gerente de Auditoria em TI do CEDASC (autarquia vinculada ao Tribunal de Contas do Estado da Bahia) e professor titular da Universidade Salvador.	26/02/2015	Entrevista por e-mail	_____
9.Luciana Sodrê Costa	Diretora do Instituto Big Data Brasil. Pós-graduada em Gestão do Conhecimento e inteligência empresarial pelo Coppe/UFRJ. Mestranda do Programa de Engenharia de Produção da Coppe/UFRJ.	27/02/2015	Entrevista realizada por Skype e telefone	40 min
10. Marcos Cavalcanti	Coordenador do Crie/UFRJ. Doutor em Informática pela Université de Paris XI, graduado em matemática pela UFRJ.	05/05/2015	Entrevista realizada por e-mail	_____
11.Sasha Issenberg	Jornalista Político. Correspondente em Washington da revista The Monacle. Tem artigos publicados New York Times Magazine, New Yourk, The Atlantic e The Washington Monthly. É autor de vários livros entre eles <i>The Victory Lab: the secret Science of Winning Elections</i> (2012)	19/05/2015	Entrevista realizada por e-mail	_____
12.Alessandro Barbosa Lima	Jornalista, especialista em Mídias Sociais. CEO da E.life. Mestre em Comunicação pela Escola de Comunicações e Artes (ECA) da Universidade de São Paulo (USP). primeira revistas eletrônica da Internet brasileira, a Revista Mundi.	20/05/2015	Entrevista realizada por e-mail	_____

Fonte: Autora

Para um melhor entendimento e visualização foram criados tópicos de apresentação dos depoimentos a partir de trechos selecionados das entrevistas, que são apresentadas a seguir.

6.4.1 Mudanças com o *big data*

Em 2008, o físico e editor da revista “Wired”, Chris Anderson, escreveu que a era *big data* representaria o fim da teoria, já que não era preciso especular sobre o melhor modelo, pois bastaria aplicar um algoritmo sobre os dados empíricos.

Previsões baseadas em dados podem ter sucesso – e elas podem falhar. É quando negamos o nosso papel no processo que as chances de fracasso aumentam. Antes de exigir mais de nossos dados, precisamos exigir mais de nós mesmos (SILVER, 2013, tradução nossa).

Essa frase de Nate Silver, estatístico norte-americano que ganhou fama depois de prever com exatidão a reeleição de Obama em 2012 é importante porque mostra como a integração de habilidades é importante.

Para o antropólogo e publicitário Renato Pereira, que têm experiências em grandes campanhas eleitorais tanto nacionais quanto internacionais, as análises de dados são importantes na condução das ações das campanhas porque ajudam aos profissionais de marketing a terem informações mais precisas sobre o processo eleitoral.

“Hoje as campanhas devem ser pensadas 95% a partir de dados estatísticos e informações precisas e apenas 5% na intuição dos profissionais. Não se concebe mais fazer uma campanha eleitoral baseada somente as experiências e intuições dos profissionais de marketing” (PEREIRA, 2014).

O publicitário ainda ressalta que o mundo passou por muitas transformações, assim como as pessoas mudaram e conseqüentemente os eleitores também.

“É preciso entender o que os eleitores acham e querem de um governo, de um político ou de um candidato. E nisso, na busca por informações e dados, o big data pode ajudar a conduzir uma campanha com mais certezas” (PEREIRA, 2014).

Da mesma forma o jornalista e especialista em mídias sociais Alessandro Lima concorda com Renato Pereira de que com o *big data* a estratégia política passa a ser menos intuitiva e mais baseada em dados. Porém, é preciso levar em conta que isso vai exigir mais atenção das campanhas e maior investimento tanto profissional quanto orçamentário.

“De uma certa forma é bom para o estrategista político pois indica mais claramente caminhos a tomar. Por outro lado exige monitorar mais de perto e criar modelos estatísticos avançados para gestão dos dados, o que acaba encarecendo campanhas. Isso é complicado pois vivemos uma época onde o investimento em campanhas está sendo duramente atacado. Mas a tecnologia fica cada vez mais barata. Hoje com ferramentas gratuitas como a que usamos já dá para armazenar e analisar muita informação” (LIMA, 2015).

Da mesma forma a presidente do Instituto Big Data Brasil (UFRJ) Luciana Sodré Costa também é otimista, dizendo que:

“O que estamos vivendo com o big data é algo sem volta. A tendência é que cada vez mais nossas vidas fiquem mais digitalizadas e dependendo dessas análises para a tomada de certas decisões” (COSTA, 2015).

Mas ela ressalta que o *big data* não substitui o ser humano no processo. Segundo Costa, ao contrário do que se imagina *big data* não pode ser pensado apenas como uma possibilidade tecnológica.

“O ser humano está presente o tempo todo no processo do big data, desde a produção dos dados até o seu estágio de análise, portanto não pode ser visto somente como uma tecnologia que vai resolver todos os problemas” (COSTA, 2015).

Costa pensa o *big data* como um ecossistema, que se alimenta e se realimenta o tempo todo com as informações, dados, ações e relações que estão sendo diariamente produzidas pelo homem em diversos ambientes do mundo digital. Segundo ela, nossas vidas estão cada vez mais sendo digitalizadas e esse processo só tende a aumentar. Isso tudo deve ser olhado por um prisma estratégico, podendo ser aproveitado o grande volume de dados digitais como uma grande oportunidade para o campo dos negócios.

“Cabem aos especialistas analisar e tirar valor estratégico para qualquer tipo de negócio. Os dados e informações que são produzidos por nós nessas relações digitais, podem ser capturados e analisados facilmente por máquinas e softwares específicos” (COSTA, 2015).

Voltando-se para o campo da política eleitoral, a cientista política e coordenadora de pesquisas eleitorais Carolina de Paula observa que o uso do *big data* pode trazer grandes transformações para uma campanha eleitoral. Ele aponta algumas dessas transformações.

“Traz racionalidade e organização. E a obrigatoriedade de investimento em profissionais que saibam articular e gerir o big data, sejam eles oriundos das áreas de TI ou aqueles que estão no dia a dia da campanha nas ruas” (PAULA, 2014).

Além disso, a cientista política diz que os candidatos que usam o *big data* em suas campanhas terão condições de economizar recursos, falando com eleitores certos.

“Ele (candidato) economizará recursos e falará somente com quem ele precisa ou deseja falar. Campanhas eleitorais são feitas em boa parte baseadas na retórica, terá vantagens quem “acertar” a retórica do momento e do seu eleitor mediano” (PAULA).

Já para Cezar Taurion (2015), usar o volume massivo de informações de uma variedade de fontes tende a diminuir os riscos de erros, quando se parte de decisões a partir do *big data*. Taurion exemplifica isso em relação às campanhas eleitorais.

“É possível com o big data fazer o cruzamento a partir das análises de nuvens de palavras. Para entender isso, basta pegar os discursos dos candidatos nos últimos tempos e comparar com o que as mídias sociais estão dizendo sobre esses candidatos e os temas. A partir daí é possível obter insights importantes para as campanhas” (TAURION, 2014).

Taurion também busca ilustrar essa parte da entrevista dizendo que o *big data* deve ser visto como um *iceberg*, onde a parte submersa são as coletas e análises de dados e a parte visível é a interface.

“Entendo que o big data começa o processo com a ciência bruta – a matemática, a estatística, as análises -, mas acaba-se com a forma mais sutil, como contar uma história, assim é preciso o trabalho dos profissionais na área de ciências sociais e humanas para tirar informações e respostas para resolver os problemas” (TAURION, 2014).

Por outro lado, o jornalista e consultor em campanhas eleitorais, Carlos Manhanelli, não acredita muito que *big data* possa ditar as regras das campanhas eleitorais. Segundo ele, a informação e os dados são importantes para o direcionamento de muitas ações de uma campanha, porém não podem ser somente isso. Segundo Manhanelli:

“Se matemática ganhasse campanhas não era preciso fazer eleições. Poderíamos pesquisar qual o candidato com mais intenções de votos e simplesmente colocá-lo no cargo. Não precisaríamos elaborar estratégias de comunicação e de marketing para convencer os eleitores de votar num determinado candidato. Acho que reduzir as campanhas eleitorais a números e estatísticas é desconsiderar o que viemos há décadas fazendo no marketing e na comunicação. É tornar a campanha fria e puramente matemática, jogando fora todos os conceitos do comportamento eleitoral e das ciências da comunicação” (MANHANELLI, 2014).

Manhanelli considera importante se utilizar técnicas de análise de grandes volumes de dados (ele não usa a palavra *big data*, prefere tratar o tema como dados) porém, diz que eles devem ser analisados também pelos olhos humanos e deve ser feito por um profissional que domine a área das ciências sociais, ciências políticas ou da comunicação.

“O computador não tem a capacidade de fazer análises subjetivas e muito menos de analisar conjunturas e nem de apontar as melhores técnicas de lançar a campanha de um candidato. Você tem informações, tem auscultas, têm sentimentos de rua dos eleitores que tem que ser levados em contas por um consultor de campanhas eleitorais e isso o computador não tem a capacidade de fazer. Então, de nada adianta um amontoado de dados e relatórios sobre cenário eleitoral e comportamento dos eleitores se o profissional, envolvido na campanha, não souber usar de forma estratégica e eficiente essas análises. As campanhas eleitorais precisam de profissionais com experiência e conhecimento da área de ciências humanas e de comunicação para dar o caminho das estratégias comunicacionais do candidato” (MANHANELLI, 2014).

A fala de Manhanelli junto a afirmação de Silver, apresentada no início desse item da pesquisa, dá o entendimento de que os profissionais de marketing e comunicação não tem o lado matemático para entender o processo do *big data*. No entanto, os analistas e cientistas de dados que, em oposição aos primeiros, sabem minerar, captar e entender os dados, são incapazes de produzir análises criativas, direcionadas e, principalmente, com alguma carga de intuição para produzir resultados superiores.

Da mesma forma, o professor e pesquisador Walter Lima Junior observa que para entender como o *big data* funciona o profissional tem que buscar um pouco de referência nas ciências duras, na matemática e segundo ele nem todos os profissionais têm conhecimentos

sobre isso, principalmente da área das ciências sociais. Na visão dele, o que torna o *big data* muitas vezes algo confuso do ponto de vista prático.

“Falta conhecimento mais aprofundado do tema. Por isso, ainda se faz tanta confusão em relação ao que big data pode fazer ou não com o grande volume de dados e informações que estão sendo produzidos pela humanidade” (LIMA JÚNIOR, 2014).

Na fala de Lima Júnior fica evidente que entre a teoria e a prática do *big data* ainda existe uma lacuna grande a ser preenchida pelo conhecimento da potencialidade e funcionalidade da ferramenta.

O ponto central dos depoimentos sobre as expectativas do *big data* é de que existe mais teorias discutidas do que práticas exploradas no que se refere a sua funcionalidade, principalmente no campo da política eleitoral. O que ainda fica difícil de ser divulgado como uma grande oportunidade e vantagens para as campanhas eleitorais, que se utilizarem do *big data* em suas estratégias de comunicação e marketing eleitoral.

6.4.2 O *big data* em campanhas eleitorais

O consultor de campanhas eleitorais Maurício Moura¹¹³ (2015) destaca que o conceito de *big data* vem sendo aplicado em campanhas políticas eleitorais no Brasil desde 2010. Segundo ele, apesar de muitos acharem que o uso do *big data* só passou a ser uma prática a partir da campanha de Obama em 2012, na realidade a política americana tem experimentado dessa prática com mais ênfase desde 2000 na campanha do presidente republicano W. Bush. E isso também é confirmado pelo jornalista americano Sasha Issenberg (2015) que conta que entre as eleições presidenciais americanas de 2000 e 2004 aconteceram inovações importantes que fizeram a coleta de dados e a análise desses dados possível.

“Os grandes saltos na arquitetura de banco de dados e no poder computacional tornaram possível gerenciar grandes listas de eleitores e executar modelos estatísticos através deles. Hoje nos Estados Unidos existem cerca de 170 milhões de pessoas registradas em bancos de dados eleitorais. E essa é uma prática feita tanto pelo partido democrata quanto pelo republicano” (ISSENBERG, 2015).

Essa prática em campanhas americanas também foi referenciada pelo diretor do Instituto Big Data da Universidade de Virgínia dos Estados Unidos, Donald Brown¹¹⁴.

“O emprego do big data em campanhas eleitorais antecede a famosa campanha de reeleição de Obama. Nos Estados Unidos os partidos e políticos se preocupam com isso há mais de uma década, principalmente no que se refere a coletar dados de eleitores e fazer cruzamentos com seus comportamentos de consumo” (BROWN, 2013)

¹¹³ Maurício Moura recebeu premiação na Pollie Award 2015, premio da AAPC (American Association of Political Consultants) na categoria internacional best in show pelo uso de WhatsApp como ferramenta de contato direto nas campanhas de 2014.

¹¹⁴ Donald Brown foi entrevistado pela autora da tese quando estava participando como palestrante no COPPE/UFRJ sobre a temática *big data*.

Para alguns especialistas, as experiências com essas práticas americanas têm influenciado também as campanhas eleitorais brasileiras.

Sobre isso, a cientista política Carolina Almeida de Paula diz que:

“Tanto o mercado quanto a academia costuma designar a primeira campanha de Obama nos EUA em 2008 como um marco num sentido de “alerta” para o uso do big data na centralidade das ações de campanha. Notamos que no pleito de 2010 no Brasil começam a aparecer debates sobre o assunto, mas o uso efetivo em especial ainda é muito pequeno neste ano. Acredito que a evolução tenha acontecido em grande medida quando as mídias sociais passam a estimular uma campanha mais pulverizada, quando notamos a horizontalidade do fluxo de informações. A tendência é que esse fator só aumente” (PAULA, 2014).

Para Renato Pereira a evolução do marketing político está no investimento maciço no conhecimento do eleitor. Segundo ele, isso é possível com o uso do *big data* e também com o técnicas de segmentação e microsegmentação, que ele chama de *microtargeting*.

“O que antes era feito somente por meio de pesquisas quantitativas e qualitativas, hoje é substituída por métodos ligados às disciplinas das ciências sociais. Isso nos leva a entender com mais profundidade o que faz o eleitor se comportar de uma determinada maneira ou tomar determinadas decisões” (PEREIRA).

A respeito disso, Pereira explica que campanhas políticas atuais têm investido em equipe de inteligência para o conhecimento do eleitorado. Segundo o publicitário, os profissionais são responsáveis por reunir dados, combinar informações e traçar perfis individuais que permitam formas de comunicação muito mais eficientes entre o candidato e os eleitores.

“Esse tem sido o objetivo do uso do big data combinado com as técnicas de microtargeting em campanhas eleitorais” (PEREIRA, 2014).

Pereira dá exemplos de como os dados são usados em campanhas eleitorais para se ter um eleitor mais segmentado.

“Você pode cruzar essas informações com as informações do PNAD, então cruzando as informações PNAD você faz uma correlação entre o lugar que a pessoa vive e conseqüentemente se ela vive em um determinado lugar ela tem um determinado perfil sócio econômico. Certo? ela tende a ser da classe E, D, C, C1, C2, B1, B2, A1, A2. Você tem o perfil sócio econômico da pessoa e o padrão presumido de consumo dessa pessoa em função do endereço dela. Se você cruza essas informações com as informações do TRE e das zonas eleitorais você sabe como, é cada uma dessas zonas eleitorais votou, então você faz um histórico delas, uma regressão no tempo de vinte anos ou vinte poucos anos, você percebe se em geral pessoas, com aquele perfil sócio econômico, morando naquele lugar, com que tipo de candidato essas pessoas costumam votar” (PEREIRA, 2014).

Assim para ele esse é o grande diferencial das campanhas eleitorais que estão utilizando o *big data* em suas estratégias de marketing eleitoral. No entanto, ele observa que no Brasil esse tipo de campanha ainda está se delineando.

“No Brasil você tem um tipo de marketing político, que normalmente que foi construído em torno da televisão, né. E, durante muito tempo os profissionais de comunicação política no Brasil deram muito pouca atenção a todas as formas de marketing direto, marketing de relacionamento e ao marketing individualizado. Por que? Porque tem o canhão da televisão, então toda atenção é dada a televisão, você não tem que comprar a mídia

que todo mundo acaba tendo mais ou menos tempo de televisão, você acaba tendo algum tempo de televisão, razoável... então por cultura ou por característica da propaganda eleitora do Brasil ou pela combinação das duas coisas a comunicação política no Brasil se orientou muito mais para o marketing televisão do que pelo marketing direto. Consequentemente toda parte de banco de dados mereceu muito menos importância e consequentemente também tudo que se refere à rede social e internet também. Nos últimos anos é que isso vem se fortalecendo, alguns poucos profissionais hoje no Brasil que ainda dão uma atenção, ou que já dão uma atenção especial a tudo isso, a maioria ainda está voltada somente para a televisão” (PEREIRA, 2012).

Já Carlos Manhanelli disse não ter trabalhado na prática com nenhuma campanha que tenha usado como referência o *big data*. Por isso, ele não consegue precisar se haveria grandes diferenças em relações as campanhas que trabalham ainda de forma mais tradicionais, usando as pesquisas de intenção de votos e pesquisas internas desenvolvidas pelas campanhas. No entanto, ele salienta que o marketing político sempre trabalhou com dados e informações de eleitores, buscando a segmentação desses eleitores. Além disso, utilizando diversos tipos de tecnologia para esse processo.

“Informação é a base para qualquer campanha, seja ela proporcional ou majoritária. O que big data promete é no meu ponto de vista, apenas mais agilidade no processo, que antes era feita somente com bases em pesquisas eleitorais ou as pesquisas internas da campanha. Hoje podemos captar informações de diversos meios, principalmente da internet para compor as informações da campanha” (MANHANELLI, 2012).

Manhanelli fala sobre sua experiência vivenciada nos Estados Unidos, nas duas últimas semanas da campanha para reeleição de Barack Obama em 2012, quando diversos especialistas em campanha eleitoral do mundo todo estiveram reunidos na University of Washington / UW. Ele salienta que lá como aqui no Brasil, a televisão é o carro chefe das campanhas eleitorais. Para Manhanelli não existe segredo na forma como se usa o *big data* e os propósitos são bem definidos nas campanhas americanas.

“O uso que eles fizeram do big data e da internet, ao contrário do que se imagina, era principalmente para arrecadar recursos para a campanha para que pudesse ser investido em comerciais políticos na televisão. Por mais interativo que seja a internet, ela não te dá votos. A televisão ainda é maior aliada de um candidato numa campanha eleitoral. E aqui no Brasil isso não é diferente” (MANHANELLI, 2014).

Sobre o uso do *big data* em campanhas eleitorais ele ainda diz que é muito mais propaganda de tecnologia do que efetivamente de uma técnica ou ferramenta revolucionária para o marketing eleitoral.

“Não vejo grande diferença não. Tudo isso, o que está acontecendo com a campanha eleitoral e a informática é o seguinte: muita propaganda. O que já apareceu de gente querendo me mostrar produtos de informática para campanha eleitoral você não tem ideia. É fila de gente querendo me mostrar. Mas é tudo igual. Eles acham que encontraram o ovo de Colombo, mas eles não acharam nada, porque isso já existe há muito tempo. Sempre tivemos a preocupação com a coleta de informações e dados sobre a situação eleitoral para dar o direcionamento das campanhas... Eu acho engraçado a internet virar onda depois da campanha eleitoral de Obama em 2008. Eu uso a internet desde 1996 em campanhas eleitorais e tenho como prova uma entrevista que dei para o jornal Gazeta, dando entrevistas de duas páginas falando de estratégias políticas, que foram instaladas aqui no Brasil com o uso da Internet. E agora todo mundo acha que a informática e big data vai resolver todos os problemas de campanha eleitoral” (MANHANELLI, 2012).

Na sua fala Manhanelli volta a reiterar que tecnologia sem um bom profissional de marketing a frente de uma campanha não ganha nenhuma eleição.

Por outro lado, César Taurion também é um entusiasta das potencialidades do uso do *big data*. Ele diz que em campanhas eleitorais o uso do *big data* poderá trazer bons resultados para os candidatos na condução das suas ações de campanha. Mas Taurion afirma que as reais potencialidades do *big data* ainda não foram totalmente conhecidas. Segundo ele o *big data* precisa ser mais explorado, experimentado em todas as direções, para que se possa dizer precisamente se é algo bom ou ruim para qualquer atividade.

“O big data é como o surgimento da internet. Ninguém conseguia prever, naquela época, o nascimento de empresas bilionárias como uma Amazon (criada justamente em 1995), de um Google (surgiu em 1998) e muito menos de um Facebook (2004), bem como as grandes mudanças que a Web provocou na sociedade. Assim é o com o big data, hoje ainda ficamos mais nas expectativas das mudanças que poderão acontecer. Mas certamente, as experiências que temos em outras áreas, e que demonstram que big data tem sido uma coisa boa para os negócios, acho que para as campanhas eleitorais, o big data poderá trazer muitos benefícios”. (TAURION, 2014).

Da mesma forma Luciana Sodr  Costa observa que o est gio em que a sociedade se encontra   irrevers vel. Segundo a especialista, o mundo digital est  presente no nosso dia a dia e n o temos mais como fugir disso.

“  preciso entender o big data como uma alternativa para entendermos e gerenciarmos o volume cada vez maior de informa es digitalizadas e as campanhas eleitorais tender o tamb m a seguir o mesmo curso, assim como em qualquer  rea do conhecimento” (COSTA).

Por outro lado Maur cio Moura diz que o importante do *big data* n o est  somente na tecnologia que se usa para se fazer essas an lises, mas sim come a na rela o da busca pela informa o e pelo dado, a partir da coleta, nas rela es estabelecidas com as fontes.

“Antes de discutir tecnologia para se implementar projetos de big data   importante identificar e maximizar as oportunidades que voc  tem com os eleitores em uma campanha. Potencializar que qualquer intera o que seja nas ruas, nas redes sociais, por telefone, nas pesquisas, sirvam de retroalimenta o dos seus dados. Esse   o principal conceito que eu procuro estabelecer nas campanhas que eu participo. De que toda a intera o   uma oportunidade de coletar dados e a partir disso   preciso estabelecer uma maneira de se organizar os dados, que seja uma maneira simples de coletar, de organizar, de alimentar as bases de dados e de se segmetat e extrair informa es a partir dessas bases de dados” (MOURA, 2015).

Pereira e Moura tamb m se mostram otimistas em rela o   tend ncia do uso de an lise de grandes volumes de dados para entender cada vez mais o comportamento do eleitor, mas ressaltam que o marketing direto e individualizado ser  um trabalho sempre paralelo ao marketing de massa, porque a televis o ainda continua sendo o carro chefe dos meios de comunica o e mais atraentes para as campanhas pol ticas. Isso tudo porque no Brasil o hor rio de propaganda pol tica   gratuito e dependendo da coliga o que o candidato

faça, sempre tem um tempo considerável para apresentar suas propostas. Segundo Pereira, o uso do *microtargeting* como método de fazer campanhas, já largamente utilizado em outros países, é o que se revela mais eficiente.

“A televisão faz um contato muito amplo com o eleitorado. O discurso generalista apresentado nos programas de TV ajuda o candidato a ficar mais conhecido, mas é o discurso voltado às questões locais que decide o voto” (PEREIRA, 2014).

Da mesma forma, o jornalista e especialista em mídias sociais Alessandro Lima (2015) acredita que futuramente o mercado brasileiro terá uma expansão no uso do *big data* em campanhas eleitorais.

“Por conta da necessidade de tomada de decisão mais embasada teremos uma expansão desse mercado. Creio que os dados de redes sociais são mais fáceis de serem coletados, armazenados e processados. Mas outros dados como o do IBGE estão aí abertos para qualquer um usá-los. Creio que haverá duas direções: conhecer o comportamento do eleitor e o próprio eleitor, avaliar os rumos de uma campanha e no futuro fazer previsões de comportamento dos eleitores (predição)” (LIMA, 2015).

Sobre as mudanças ocorridas na comunicação eleitoral com o uso do *big data*, a cientista política Carolina de Paula diz:

“O big data atua num sentido de aproximar o eleitor da campanha, ou seja, ele vai ouvir e ver o que lhe diz respeito e lhe interessa. A comunicação política muda num sentido de ser mais geral para ser cada vez mais específica. Diante disso, você precisa de instrumentos técnicos pra isso, como o uso do Whatsapp, por exemplo, que vimos fortemente sendo utilizados por campanhas eleitorais em 2014” (PAULA, 2015).

Para o professor da COPPE/UFRJ Marcos Cavalcanti, o *big data* vai por um lado permitir que políticos conheçam melhor a diversidade de seus eleitores, para que possam estabelecer uma relação de maior qualidade e valor com eles.

6.4.3 Estrutura e especialidades profissionais

Para muitos especialistas, um dos grandes desafios no Brasil é a falta de profissionais qualificados na funcionalidade do *big data*. Um relatório do IDC do Gartner aponta que demanda por profissionais orientados a projetos de *big data* criará em 2015 em torno de 4,4 milhões de postos de trabalho no mundo. Isso ocorre porque as funções requerem habilidades que vão além do uso de painéis para monitorar o fluxo de informações, entender de tecnologia e de estratégias de mercado a partir das possíveis análises de dados oferecidas pelo *big data*. É fundamental ter experiência para ajustar corretamente o que é relevante e definir parâmetros para estabelecer filtros e algoritmos. Esse conhecimento dificilmente se atinge com meras

certificações profissionais. Muitas vezes, a exigência chega em níveis de mestrado ou doutorado.

Outro estudo¹¹⁵ realizado em 2013 pelo Instituto Brasileiro de Inteligência de Mercado (Ibramec) com 326 empresas no Brasil apontava que as técnicas de *big data* deveriam crescer 65% nos departamentos de marketing dessas corporações e 16% na área de Tecnologia da Informação.

A demanda de mercado contempla profissionais qualificados para decisões técnicas e humanas. Esse profissional, que vem sendo chamado de “cientistas de dados” (do inglês *data scientists*), cientista de *big data* ou *Chief Data Officer* (CDO) precisa ter raciocínio lógico apurado, conhecimento profundo de estatística e *software* dessa área, modelagem, domínio de sistemas computacionais e conhecimento do negócio e do mercado em que atuará. Além disso, é necessária a capacidade de comunicar seus *insights* a um pessoal com uma formação não técnica (DAVENPORT; BARTH; BEAN, 2012) voltada principalmente para a área de comunicação.

Para os entrevistados na pesquisa, o mercado de *big data* se mostra em desenvolvimento, assim como a capacitação de profissionais, principalmente da área de tecnologia da informação. Segundo Maurício Moura (2015), não existe ainda um profissional específico que seja capaz de entender todo o processo e dinâmica do *big data*, principalmente no Brasil. A saída para atender esse mercado é trabalhar com profissionais multidisciplinares.

“Na minha empresa trabalhamos com vários tipos de profissionais, que vai desde cientistas sociais e políticos, profissionais da área de tecnologia da informação e estatística até profissionais de marketing e comunicação. Porque entendemos que um profissional mesmo que tenha as qualidades exigidas de um cientista de dados, não daria conta de todo o processo e entendimento do conceito e aplicação do big data”. (MOURA).

Da mesma forma o professor Marcos Cavalcanti acredita que o mercado vai demandar de um profissional para entender de *big data* que tenha visões tanto para a área técnica quanto de negócios. Segundo ele:

“A digitalização crescente das relações sociais, da economia e da vida em geral terá forte impacto nos modelos de negócios das empresas e vai demandar um novo tipo de profissional capaz de entender de tecnologia e de negócio, com uma formação mais horizontal e sistêmica” (CAVALCANTI).

Na fala de Maurício observa-se que o mercado do *big data* está se delineando. O Brasil, ainda, não tem mão de obra específica para oferecer nessa área. Da mesma forma a

¹¹⁵ Informações retiradas do jornal Gazeta do Povo (22/02/2014).

cientista política Carolina de Paula¹¹⁶ diz que quanto mais multidisciplinar foi a equipe maior o sucesso da campanha.

“Você precisa que o cara da tecnologia escute o publicitário e que este escute o analista de pesquisas de opinião por exemplo. No fundo é isso, as campanhas hoje precisam lidar com a variedade de públicos e targets, e pra isso acontecer somente uma equipe de formação variada consegue ser bem sucedida” (PAULA).

Assim, os profissionais que já trabalhavam com dados, informações, pesquisas e estatísticas vão tentando se ajustar a essa necessidade. Para os especialistas a procura das empresas hoje é maior do que a disponibilidade de profissionais. As empresas acabam formando seus próprios talentos. Trazem alguns com algumas competências e vão desenvolvendo as demais.

Para a presidente do Instituto Big Data Brasil Luciana Sodré Costa, o grande problema hoje está não somente na falta de profissionais específicos para trabalhar com *big data*, mas no entendimento errado que as empresas e organizações têm sobre o conceito. Segundo ela, *big data* não é somente tecnologia, diz respeito ao campo estratégico e de negócios. Ela fala da sua experiência como pesquisadora do Crie/UFRJ e da falta de entendimento por parte das empresas que procuram por projetos de *big data* acreditando no poder de sanar todos os problemas de suas empresas.

“Muitas vezes recebemos aqui no Crie empresas e organizações interessadas em nossos projetos e pedem logo para adquirirem o big data, achando que somente a tecnologia, a ferramenta dará conta dos problemas de excesso de dados da sua empresa. E na verdade não é só adquirir tecnologia. O big data requer muito mais do que isso. É preciso ter profissionais disponíveis e interessados em saber como usar o big data para tirar valor da sua aplicação e reverter em resultados positivos para as organizações” (COSTA, 2015).

As observações da pesquisadora são discutidas também por outros especialistas que afirmam que as empresas ainda não se deram conta de como esse conceito pode trazer algum benefício para os seus negócios.

Compartilhando da mesma opinião de alguns entrevistados, o especialista em tecnologia da informação, Christian Guerreiro diz que o Brasil ainda está engatinhando no que se refere ao *big data*.

“Só agora as instituições de ensino começam a ofertar cursos especializados na área, refletindo a demanda tardia por conhecimento na área”(GUERREIRO, 2014).

Segundo o especialista o perfil do profissional deve contemplar conhecimentos nas áreas de TI e estatística. Mas ele ressalta que a formação de equipes multidisciplinares é

¹¹⁶ A cientista política Carolina de Paula é uma das profissionais do Ideia Inteligencia.

importante para agregar especialistas em infraestrutura, banco de dados e análise de dados para que se obtenham melhores resultados para a organização. Guerreiro ainda completa:

“Como visão de futuro, percebo que as complexidades enfrentadas hoje serão amenizadas no futuro com a disseminação da tecnologia e evolução das soluções atualmente disponíveis, que serão incorporadas ao dia a dia das empresas gradativamente, como toda inovação tecnológica” (GUERREIRO, 2014).

Guerreiro em sua fala acredita que futuramente o mercado poderá contar com profissionais específicos para atender aos projetos de *big data*, mas tudo depende de como as instituições de ensino irão promover e disponibilizar conhecimentos nessa área.

Para Cezar Taurion existem diversas formas para se trabalhar com as atividades de *big data* em campanhas eleitorais e existe lugar para diversos tipos de profissionais e até mesmo empresas.

“Existem diversas formas de se oferecerem serviços para os candidatos hoje. Se você tem uma empresa que faz monitoramento de mídias digitais e a partir disso faz análise de sentimento sobre campanhas e candidatos, é uma forma de se oferecerem serviços de big data. Esse é um negócio interessante porque você terá alguém que poderá orientar a campanha nas mídias sociais a partir das análises de conteúdos positivos ou negativos. Então, não precisa ser alguém especificamente com conhecimentos em big data e sim em análises de mídias sociais” (TAURION, 2014).

Os tipos de serviços citados por Taurion já existem no Brasil. Esse é o exemplo da empresa de gestão de relacionamento em redes sociais ELife, que foi anteriormente citada na pesquisa e que oferece assessoria para políticos e candidatos e para os meios de comunicação desde 2013, usando o conceito de *big data* eleitoral. Assim como, a própria empresa Ideia Inteligência (que também já foi pesquisa anteriormente nesse trabalho) que oferece diversos tipos de serviços no segmento político e governamental, com foco principalmente em projetos de *big data*.

6.4.4 Usos de dados pessoais

A massificação do *big data*, no entanto, ainda enfrenta obstáculos, segundo a visão de alguns especialistas. O mais comentado desses obstáculos é a questão da privacidade e ética do uso de dados, principalmente as informações de caráter pessoal. A respeito disso, na pesquisa abordou junto aos entrevistados sobre como lidavam com a questão da privacidade representada pelo aumento do armazenamento e integração de informações pessoalmente identificáveis¹¹⁷.

¹¹⁷ Informações pessoalmente identificáveis refere-se a informações que podem ser usados para identificar, contactar ou localizar uma única pessoa ou pode ser usado com outras fontes para identificar um único indivíduo.

Do ponto de vista científico, Walter Teixeira Lima Júnior¹¹⁸ (2014) observa que a ética é ponto crucial no uso dos dados para pesquisas em *big data*. Segundo ele:

“O big data para ser válido dentro da ciência tem que ser open (aberto). Tem que dar possibilidades para que as outras pessoas interessadas em pesquisas com determinados dados possam verificar como estão sendo utilizadas as metodologias em torno da pesquisa. Quem usa o big data na ciência tem que comprovar que os dados são verdadeiros do ponto de vista científico” (LIMA JÚNIOR, 2014).

No entanto, o pesquisador observa que, do ponto de vista comercial, isso é um tanto complicado e não funciona assim de forma tão transparente em se tratando de dados principalmente pessoais.

“Se uma empresa abrir os dados, outra empresa concorrente pode copiar os seus dados e tomar o mercado a partir do que essas informações dizem ou projetam” (LIMA JÚNIOR, 2014).

Assim segundo Lima Júnior, as empresas e mesmo campanhas eleitorais sempre trabalham mais no anonimato das suas ações de coleta, análise e uso de dados no desenvolvimento de estratégias e inteligência competitivas.

Da mesma forma Luciana Costa observa que o ideal seria que dados pessoais, como por exemplo, de redes sociais fossem *open* (abertos), principalmente para serem utilizados em pesquisas acadêmicas, e não, somente, para fins comerciais como as provedoras dos sites de relacionamentos usam.

“Hoje quem tem acesso total aos dados pessoais de usuários do Facebook é somente a empresa do site, que diz usá-los para melhorar a qualidade de seus serviços, mas ninguém pode garantir que isso realmente aconteça. E muitas vezes os usos que a provedora faz desses dados não são claros. Seria muito bom se pudéssemos ter acesso a esses dados para que outras pesquisas fossem realizadas, com outros olhares” (COSTA).

No que se refere a prática de usos de dados para outras finalidades, Costa ainda observa que hoje as pessoas que usam da tecnologia como redes sociais e aplicativos não têm como fugir dessa coleta de dados pessoais feitas por essas empresas.

“Ao acessar um aplicativo que te oferece serviços, você autoriza, mesmo que você não saiba, o uso dos seus dados. Então, todos os aplicativos vão te oferecer soluções para as suas necessidades, em contrapartida, você dará a eles por esses serviços seus dados que serão automaticamente coletados, analisados e usados. Mas o que se espera é que esses dados sejam utilizados de maneira anônima, porque na verdade o que interessa é essa dinâmica do comportamento humano” (COSTA, 2015).

Assim, manter a privacidade ou mesmo o anonimato diante dessas empresas é praticamente impossível. No entanto, Costa diz que é importante que os usuários tenham educação e orientações para os usos das mídias sociais, no sentido, de ao acessarem os

¹¹⁸ Walter Lima Júnior é um defensor do *open data* e da *open science*, por isso todas as suas pesquisas são publicadas em plataformas em Apis (*Application Programming Interface*) para que qualquer pessoa interessada tenha acesso ao seus bancos de dados e possam verificar como foram utilizados aquelas informações dentro de uma metodologia específica.

serviços ou fazerem adesões, possam avaliar se os serviços que querem dessas mídias e ferramentas valem os seus dados pessoais, que serão entregues e fornecidos a essas empresas.

Pensando assim, os dados e informações pessoais se configuram mais numa decisão individual a ser tomada do que propriamente de um caso de ética do uso pelas empresas que se apropriam e se beneficiam desses dados.

Cezar Taurion (2014) também chama atenção sobre as questões de privacidade dos dados. Sobre isso afirma que:

“Pessoalmente, opto pelo viés positivo, mas com fortes alertas às questões como privacidade e uso indevido e não autorizado de informações pessoais” (TAURION, 2014).

Segundo ele um dos grandes desafios que precisa ser debatido com relação ao *big data* relaciona-se justamente com a questão da privacidade dos usuários das redes sociais, porque somos condicionados a utilizar esses ambientes de forma gratuita na base da troca de nossas informações. É o caso do Google e Facebook que coletam essas informações, alegando em seus termos de uso e nas políticas de privacidade respeitarem veementemente a privacidade e segurança de seus usuários. No entanto, ocultam os reais interesses na utilização desses dados, que podem ser facilmente comercializados.

O Google e o Facebook, citados por Taurion, utilizam-se de ferramentas, como os *cookies*¹¹⁹, para coletar e armazenar informações pessoais. Eles afirmam em seus termos de usos do site que serão para fins de segurança, aprimoramento de produtos e serviços, e também para um melhor desempenho no uso do serviço.

Assim, essas empresas agregam informações pessoais dos usuários, tornando-se assim habilitados a usar essas informações para personalizar o que será enviado aos mesmos, como por exemplos anúncios e propagandas do interesse do seu perfil.

O jornalista americano Sasha Issenberg (2015) em sua fala para a pesquisa faz referências aos Estados Unidos, ressaltando que o uso de dados pessoais, principalmente de redes sociais são considerados públicos lá, porque as empresas de relacionamento como o Facebook deixam isso claro para os seus usuários no momento em que eles aceitam os seus serviços. Por outro lado, a própria cultura americana é baseada na disponibilização e divulgação de dados pessoais, principalmente para o governo. Da mesma forma, os partidos (Republicano e Democrata) buscam manter atualizados suas bases de dados com informações pessoais e precisas sobre os americanos. A preocupação do governo americano em ter dados

¹¹⁹ Segundo a definição de Sawaya (2010), *cookie* é um arquivo usado pelos servidores para manter rastro dos padrões e preferencias dos usuários quando este entra em certas páginas da web.

dos americanos a sua disposição se intensificou a partir dos atentados de 11 de setembro de 2001, a fim de manter controle sobre as ações dos cidadãos.

Compartilhando da mesma opinião de Sasha, o jornalista Alessandro Lima ressalta sobre os usos dos dados pessoais das redes sociais digitais.

“Os dados são abertos e os eleitores autorizaram o seu uso uma vez que concordam com as políticas de privacidade das principais redes sociais. As redes por sua vez apenas comercializam estes dados para empresas outras empresas. E muitas vezes os dados são anonimizados. Por isso, não vejo muitos problemas em relação a isso desde que essas informações e dados sejam usadas de uma forma positiva em relação aos usuários dessas redes” (LIMA).

Já para a cientista política Carolina de Paula, o uso de dados pessoais também se mostra como um problema em relação as campanhas eleitorais. Porém, nem sempre é possível coletar dados tão precisos assim para o *big data*. Isso deve-se ao fato de que existe leis que protegem os dados pessoais.

“Esse é o ponto mais delicado sobre o uso de big data, sem dúvida. Por isso temos leis que protegem severamente a privacidade do indivíduo. Sabemos que há possibilidades técnicas de obtenção de dados melhores, via hackers, por exemplo. Mas o profissional precisa saber que este não é um jogo de vale tudo, né? Sempre teremos dados “piores” do que aquilo que potencialmente poderíamos usar devido à proteção do Estado, mas é bom que isso continue assim” (PAULA, 2014).

Da mesma forma o professor Marcos Cavalcanti defende que:

“O que deve ser preservado é a anonimização dos dados. Acredito que o processo de geração e compartilhamento de dados é um processo irreversível e que tende a aumentar. O que devíamos assegurar, além da anonimização é que os cidadãos terão acesso a seus dados e que eles não poderão se tornar propriedade de nenhum governo ou empresa” (CAVALCANTI).

Na visão geral dos especialistas, os dados pessoais sempre serão uma questão a ser debatida, porém acham que seja provável que soluções mais eficientes das que já existam garantidas por lei para manter o direito aos dados pessoais e o anonimato a curto prazo devam aparecer, já que o uso do *big data* em diversos setores está se expandido.

No que se refere as campanhas eleitorais isso também vem sendo visto como um problema, já que candidatos e campanhas se utilizam de todas as estratégias na busca por votos, principalmente diante de uma disputa acirrada.

6.4.5 Pontos negativos do *big data*

Para o consultor em campanhas eleitorais Maurício Moura a maior dificuldade seria fazer as campanhas entenderem da necessidade de se trabalhar ações com bases em análises de dados, prevendo o comportamento eleitoral. Essa dificuldade segundo ele é sentida ainda aqui no Brasil, em que as campanhas eleitorais seguem uma hierarquia de cima para baixo, principalmente em se tratando de campanhas marjoritárias (governador, prefeito e presidente). Na fala dele isso fica mais claro:

“No Brasil a coleta de dados é incipiente, mas quando se começa a fazer é um caminho sem volta. Você vê a importância e você vê o resultado dessa utilização. Mas depende de disciplina, ainda mais no caso das campanhas eleitorais que as estruturas são de cima para baixo. Assim, o big data ainda está muito na base dos seus direcionamentos. Muitos ainda não entendem a sua real importância principalmente em campanhas eleitorais em que a disputa é apertada, porque cada dado, cada informação pode fazer a diferença e se reverter em voto para o candidato.” (MOURA, 2015).

Já para a cientista política Carolina de Paula (2014) a maior dificuldade de se trabalhar com o *big data* é a desorganização dos dados e informações. Ela exemplifica isso, a partir de uma campanha de reeleição de um determinado governante. Muitos dados coletados pelo governo nem sempre estão disponíveis e organizados para uso imediato, nem mesmo para posicionar a campanha em relação as várias áreas administrativas como saúde, educação, segurança. As informações obtidas dessas áreas poderiam ajudar a montar o perfil daquele político ou campanha, estão sempre descentralizados, espalhados.

“Os dados que existem estão em bases e bancos de dados descentralizados. Os dados são muitos desorganizados ou inexistentes. Falta integração dos dados entre as secretarias e órgãos administrativos. Falta conhecimento mais apurado das características das suas regiões, por exemplo. Isso é possível ver principalmente em épocas de campanhas eleitorais, quando precisamos de dados e informações para compor as estratégias de pesquisa por regiões. E isso faz com que percamos muito tempo coletando dados, que já foram coletados em algum momento pela aquela secretaria. Assim, esse trabalho acaba sendo refeito” (PAULA, 2014, 2015).

Outro ponto negativo a ser considerado nos usos do *big data* em campanhas eleitorais brasileiras, se referem às questões orçamentárias e disponibilidade financeira das campanhas em investir em profissionais especializados em coleta e análise de grandes volumes de dados. Segundo Moura (2015) a exatidão das análises dos dados eleitorais necessita de dados precisos e que muitas vezes acabam por encarecer o processo, principalmente de coleta.

“A era analógica de coleta e análise de dados era muito cara e demorada, mas com a tecnologia o processo de coleta ficou mais ágil, mais barata, mas existem alguns dados que é preciso fazer a coleta na fonte, porta a porta. Esse processo é demorado, demanda metodologia de pesquisa e assim precisamos de profissionais qualificados para cuidar de todo o processo” (MOURA, 2015).

Ainda sobre os pontos negativos do *big data*, Moura complementa a sua fala dizendo:

“O maior risco de se usar o big data em uma campanha eleitoral é não ter um diagnóstico adequado. O melhor diagnóstico passa por uma análise quantitativa de dados. Portanto, não vejo como ter um diagnóstico de uma campanha eleitoral sem ter uma análise quantitativa desses dados. A partir daí é que podemos empreender uma análise mais qualitativa”. (MOURA, 2015)

Da mesma forma Renato Pereira ressalta:

“Acho que um dos pontos negativos é a própria cultura da política brasileira de não se preocupar com que os eleitores estão dizendo e querendo dos candidatos, dos governantes. É preciso direcionar a campanha para o lado certo, que vai realmente trazer resultados positivos para o candidato. Mas isso demanda entendimento e valorização do processo. Por outro lado, o big data não é uma realidade a todas as campanhas brasileiras. Porque não se trata apenas de ter tecnologia disponível, é preciso investir em profissionais e ter disposição e orçamento para isso” (PEREIRA, 2014).

No que se refere aos estudos mais aprofundados sobre os eleitores brasileiros e seus comportamentos em relação à política, Moura também enfatiza que o Brasil ainda está distante de ter uma “microvisão” estatística apurada de sua população.

“As bases de dados públicas disponíveis por CPF apresentam pouquíssimas informações e podem ser consideradas pré-históricas quando comparadas aos dados disponíveis nos Estados Unidos, França ou Inglaterra. O que dificulta muitas vezes o trabalho a partir desse tipo de coleta de informações. Aí temos a necessidade de um trabalho de formiguinha, de levantar informações porta a porta, diretamente na fonte. Por isso hoje a importância de termos uma empresa de dados no Brasil voltada para isso” (MOURA, 2015).

Na opinião de Carlos Manhanelli o ponto negativo do uso do *big data* é achar que somente a tecnologia poderá dar conta de ganhar uma campanha eleitoral.

“O cruzamento de dados só é útil se ele tiver a análise apurada de um profissional com experiência e talento para entender do que as campanhas eleitorais precisam. Precisamos de profissionais específicos para tomar as decisões. Não é o computador que vai decidir como isso deverá ser feito, nem tão pouco um cara da informática. Esse é o grande erro da expectativa com essa terminologia do big data, em achar que é a solução para a conquista de uma disputa eleitoral e que as experiências dos profissionais e consultores de marketing eleitoral podem ser dispensadas a partir de análise de grande volume de dados” (MANHANELLI, 2014).

Manhanelli ainda sobre isso complementa, dizendo que as promessas do *big data*, sempre foram de certa maneira realizadas pelas consultorias de marketing eleitoral.

“Hoje temos mais dados disponíveis em diferentes formatos e lugares que podem nos ajudar a entender melhor o comportamento eleitoral, mas isso não quer dizer se eu não adotar o big data, minha campanha estará fadada ao insucesso. É preciso avaliar em que situação é importante o seu uso. Não vejo a necessidade em tornar isso tudo como um padrão no marketing eleitoral” (MANHANELLI, 2014).

Da mesma forma o professor Marcos Calvacanti aponta que a maior desvantagem em relação ao *big data* é as pessoas acharem que ele é uma tecnologia e que por si só vai resolver problemas, que hoje não sabemos como resolver.

“O grande erro de adotar o big data é achar que é uma panaceia que vai resolver tudo, quando na verdade não vai. Ele apenas vai responder as perguntas que lhe foram feitas” (CAVALCANTI, 2014).

Ele ainda diz:

“Se todas as respostas estão na internet, qual o papel restou para os seres humanos? Fazer as perguntas... O big data não pode fazer perguntas. Ele pode nos ajudar a encontrar relações entre dados que desconhecíamos, mas continuamos a depender da inteligência e sensibilidades humanas para colocar as grandes questões que nos fazem progredir” (CAVALCANTI, 2014).

Para Cavalcanti a exclusão do ser humano do processo do *big data* se torna o grande ponto negativo de se pensar uma sociedade cada vez mais digitalizada.

6.5 CONSIDERAÇÕES DA PESQUISA DE CAMPO

Nas entrevistas realizadas com os especialistas foi possível observar que, tanto do ponto de vista prático quanto do ponto de vista científico e acadêmico, o *big data* se coloca como um grande desafio para as campanhas eleitorais e para as organizações. O maior problema levantado pelos especialistas é fazer com que as pessoas entendam qual é a real função do *big data*, seja numa campanha eleitoral ou em áreas de negócios.

Nas suas falas, todos colocam que o *big data* não pode ser entendido apenas como uma tecnologia e não se pode excluir a relação humana de todo o processo, desde a geração de dados até a sua análise e uso. Assim um dos maiores desafios do *big data* na visão desses especialistas é perceber não como uma solução em tecnologia da informação, mas como um ativo que deve ser acionado para encontrar as melhores soluções dentro do planejamento estratégico de uma campanha eleitoral ou de qualquer outro tipo de negócio.

A coleta dos dados pessoais e usos desses dados se coloca como um ponto negativo do *big data*, além da falta de entendimento o que seja o *big data* e de como ele pode ser usado para ajudar a resolver problemas tanto de campanha quanto de empresas. Quando Moura diz que no Brasil o sistema de coleta de informações pelo CPF da pessoa é praticamente inexistente, isso de certa forma contradiz o que muitos defendem em relação ao uso de dados pessoais. Entende-se que a anonimização se refere à exclusão de quaisquer dados pessoais de um banco de dados, como nome, endereço, número de cartão de crédito, data de nascimento ou identidade. Isso pode funcionar como diz Schönberger-Mayer e Cukier (2013) para os pequenos bancos de dados, mas em grandes bancos de dados, como é o caso do *big data*, que com o aumento e variedade de informações, facilitam a reidentificação. Assim disponibilização de dados específicos via o CPF, cruzados com outros dados vindos inclusive de redes sociais de uma pessoa pode simplesmente identificá-la e isso se configura como um dos grandes debates sobre o *big data* na atualidade.

Então, se por um lado se defende o anonimato nessas coletas, então é bem melhor que esses dados vindos mesmos de listas consideradas “públicas” continuem a apresentar problemas de falta de informação, como bem pontua a cientista política Carolina de Paula. Já no que se refere às expectativas em relação ao uso do *big data* em campanhas eleitorais, percebe-se que por mais otimistas que os entrevistados sejam em relação o seu uso no Brasil, ainda se está engatinhando no processo. Para os especialistas em campanhas com o uso do *big data* Maurício Moura e Renato Pereira, as campanhas ainda precisam mudar muito a sua forma de pensar a disputa eleitoral. Segundo eles, o marketing eleitoral no Brasil se concentra muito na disputa na televisão e usa muito pouco as outras formas de marketing como o

marketing direto, aquele feito porta a porta, de forma individualizado. Assim, além de contar com estruturas e orçamento para investir em uma campanha centrada em dados, seria necessário começar a mudar a cultura das campanhas, que são mais voltadas para a massificação da comunicação do que para a segmentação das mensagens.

Outro ponto da pesquisa, muitos entrevistados colocam o *big data* como uma transformação na forma de se fazer campanha. Isso, no sentido de que a partir das análises de grandes volumes de dados, poderá se tirar informações mais precisas para que se conduzam campanhas com mais clareza e precisão, demandando menos gastos em comunicação eleitoral com pessoas (eleitores) que não sejam potenciais para o candidato, ou mesmo, contribuindo para aproximar o candidato de seus eleitores, numa comunicação direta e segmentada. Por outro lado, se coloca o problema de trabalhar diagnósticos precisos a partir de análises quantitativas, que demandam tempo e orçamento para investir em profissionais qualificados desde a coleta até a análise desses dados.

E por fim, o que se conclui dessa pesquisa a partir das entrevistas com os especialistas é que o *big data* é um recurso e um instrumento que existe para informar, não explicar e que pode ajudar a direcionar as ações de marketing de uma campanha política. Assim, também entende-se que o *big data* não é um mundo frio, de algoritmos e computadores. Há um papel essencial para as pessoas no processo desde a coleta até as decisões a serem tomadas com os resultados das análises desses dados. Portanto, *big data* não substitui a figura do especialista e nem tão pouco é capaz de fazer milagres. Ainda se precisa muito da expertise dos profissionais de marketing para conduzir as melhores ações para a campanha, a partir de decisões tomadas em cima de dados e pesquisas que privilegie a totalidade de eleitores e não somente uma forma amostral como eram feitas no passado.

6.6 USOS DO *BIG DATA* NA CAMPANHA ELEITORAL DE 2014

Para realizar esta pesquisa foi necessário a integração da pesquisadora em uma campanha eleitoral na disputa eleitoral de 2014. A campanha eleitoral escolhida foi para candidato ao governo do Estado do Rio de Janeiro. A escolha por essa campanha se deu pela proximidade geográfica, pela facilidade na integração à equipe de profissionais de comunicação e marketing, e por se tratar de uma campanha que tinha como um dos focos o uso do *big data* em suas ações.

Na pesquisa optou-se por não identificar o candidato da qual a campanha era direcionada. Ele será chamado e referenciado como **candidato X**. A pesquisa realizada não

tem como objetivo analisar resultados do processo eleitoral e nem conjuntura política, apenas entender como se dá a dinâmica das ações de informação e comunicação, a partir dos usos do *big data*. Além disso, a pesquisa não procura se aprofundar nas técnicas e métodos utilizados para se chegar ao entendimento do comportamento eleitoral e nem tão pouco analisa o resultado final das eleições. Aqui procura-se descrever de que forma os dados e informações da campanha foram coletados e usados. As análises passavam pelo processo tecnológico em que *softwares* específicos e programas de *big data* após o cadastramento desses dados e sob comando da área de Tecnologia da Informação (TI), eram agrupados, cruzados e analisados de acordo com os objetivos e focos da campanha.

Para entender de forma prática como o *big data* era usado, durante os meses de agosto, setembro e outubro de 2014 foram acompanhadas as atividades do comitê eleitoral usando como metodologia de pesquisa a técnica de observação direta. Durante o acompanhamento das ações de comunicação e marketing foram também realizadas entrevistas não-estruturadas com alguns profissionais envolvidos diretamente na campanha. Esses profissionais não foram identificados nominalmente na tese. Manteve-se o anonimato de seus nomes, a fim de respeitar os seus termos contratuais com o cliente (partido). Assim, optou-se por identificá-los pelas suas especialidades exercidas na campanha, como mostra o Quadro 14.

Quadro 14: Profissionais entrevistados no Comitê Eleitoral

Código	Especialidade
AMD	Analista de Mídia Digital
CID	Coordenador de Inteligência de dados
CGC	Coordenador Geral de Comunicação
CMD	Coordenador de Mídias Digitais
CR	Coordenador de Rua
DCD	Diretor da Campanha Digital
GP	Gerente de Pesquisa
JOR	Jornalista

Fonte: Autora

6.6.1 Cenário eleitoral para governo do Rio de Janeiro

A disputa eleitoral para governo do Rio de Janeiro em 2014 girava em torno de cinco principais candidatos. Concorriam ao Palácio Guanabara o governador Luiz Fernando Pezão (PMDB), o ex-governador Antony Garotinho (PR), o senador Marcelo Crivella (PRN), o senador Lindenberg Farias (PT) e o professor Tarcísio Motta (PSOL). Além de outros candidatos como: Dayse Oliveira do PSTU e Ney Nunes do PCB, que quase não apareciam nas pesquisas de intenções de votos.

Quadro 15: Cenário Eleitoral para Governo do Rio de Janeiro em 2014

Candidatos a governador e a vice ↕	Número Eleitoral ↕	Coligação ↕	Tempo de horário eleitoral ↕	Votos no 1º turno ↕	Porcentagem ↕	Votos no 2º turno ↕	Porcentagem ↕
 Luiz Fernando Pezão (PMDB) Francisco Dornelles (PP)	15	<i>O Rio em Primeiro Lugar</i> PMDB, PP, DEM, PSD, PSC, PSDB, PPS, PSL, SD, PHS, PTN, PSDC, PMN, PTC, PRP, PRTB, PPL, PEN e PTB	8'57	3.242.513	40,57%	4.343.298	55,78%
 Marcelo Crivella (PRB) General Abreu (PRB)	10	<i>sem coligação</i>	1'09	1.619.165	20,26%	3.442.713	44,22%
 Anthony Garotinho (PR) Márcio Garcia (PR)	22	<i>Aliança Republicana e Trabalhista</i> PR, PROS, PTdoB	2'17	1.576.511	19,73%		
 Lindberg Farias (PT) Roberto Rocco (PV)	13	<i>Frente Popular</i> PT, PV, PSB, PCdoB	4'38	798.897	10,00%		
 Tarcísio Motta (PSOL) Renato do Quilombo (PSOL)	50	<i>sem coligação</i>	1'01	712.734	8,92%		
 Dayse Oliveira (PSTU) Marília Macedo (PSTU)	16	<i>sem coligação</i>	0'57	33.442	0,42%		
 Ney Nunes (PCB) Heitor César (PCB)	21	<i>sem coligação</i>	0'57	8.950	0,11%		

Fonte: Wikipédia 2014

Como mostra o quadro 15 a campanha para governador no Rio de Janeiro foi disputada em dois turnos. No primeiro turno Luiz Fernando Pezão ficou em primeiro lugar com 40,57% dos votos, em segundo lugar Marcelo Crivella (PRB) com 20,26% e em terceiro lugar Antony Garotinho (PR) com 19,73% dos votos. O segundo turno foi disputado entre Pezão e Crivella. Pezão disputou o segundo turno sem nenhum apoio dos candidatos derrotados, enquanto Crivella foi apoiado por Anthony Garotinho e Lindberg Farias (PT).

O resultado final dá vitória para Luiz Fernando Pezão (PMDB) que foi eleito com 4.343.298 votos representando 55,78 % do total. Marcelo Crivella do PRB ficou em segundo lugar com 44,22% dos votos.

6.7 DE DENTRO DO COMITÊ ELEITORAL

O comitê eleitoral ficava em um bairro de classe média alta, na zona oeste do Rio de Janeiro, para chegar ao local era preciso pegar um metrô saindo da estação Botafogo (Zona Sul) até a estação Maria da Graça no Bairro Cachambi (Zona Oeste), depois mais um ônibus, saindo do referido bairro até o local das instalações do comitê, eram gastos nesse trajeto

aproximadamente duas horas de transporte público e uma hora e meia de carro. O comitê começava suas atividades a partir das 10 horas da manhã e terminava por volta da meia noite.

O comitê funcionava durante o período eleitoral, praticamente 24 horas por dia em todos os dias da semana. Os profissionais que ocupavam funções mais de coordenações se revezavam em plantões nos finais de semana e feriados, principalmente a equipe de dados e a equipe responsável pela campanha digital, a equipe operacional fica disponível todos os dias da semana no período entre 8 a 10 horas de trabalho. Em noites que o candidato tinha aparições no Horário Gratuito da Política Eleitoral (HGPE) e participações em debates televisivos, os profissionais centralizavam suas atenções para ações de monitoramento das mídias sociais, com elaboração de relatórios em tempo real e análises de dados da campanha. Nesses dias os profissionais ficavam no comitê até meia noite, como será descrito mais adiante nesta pesquisa.

O local onde ficava o comitê de campanha era um amplo espaço (como se fosse um barracão), sem janelas, com apenas duas portas, uma entrada principal, onde ficavam duas recepcionistas e seguranças, e uma porta com saída de emergência. No centro desse espaço ficava uma pequena sala de estar e a mesa do coordenador geral da campanha, de onde conseguia ter uma visão panorâmica das atividades que estavam sendo desenvolvidas pelos demais profissionais.

No barracão existiam poucos espaços com divisórias: uma sala reservada logo na entrada, em que ficava as pessoas do partido e alguns assessores diretos do candidato (as atividades eram mais administrativas, logísticas e orçamentárias), uma sala de reuniões improvisada com divisórias móveis, onde os profissionais se reuniam muitas vezes com o gerente geral da campanha para alinhamento das estratégias de comunicação. A ideia de reunir todos os profissionais e áreas da campanha num único ambiente era de buscar uma integração entre as equipes e ao mesmo tempo fazer com que todos tivessem conhecimento do que cada área específica da campanha estava fazendo. Assim todo mundo sabia de todo mundo, mostrando uma equipe coesa e integrada com conhecimentos amplos sobre as diversas áreas, porém não perdendo a centralidade que deveria ser desenvolvida em cada atividade.

No espaço era comum ver a circulação do candidato em determinados dias da semana, principalmente, para participar de reuniões, acompanhar de perto as atividades da equipe de comunicação e marketing, bem como para avaliar suas peças publicitárias que iriam para o ar na televisão e no rádio, ou mesmo gravar programas eleitorais.

Imagem 3: Estrutura física do Comitê de Campanha



Fonte: Autora¹²⁰

A Imagem 3 demonstra como era a disposição da equipe, que trabalhava de forma integrada num único espaço chamado de Pólo. Esse local era considerado o centro de inteligência de informação e comunicação da campanha eleitoral. Desse comitê saíam todas as diretrizes de comunicação do candidato para os diversos meios de comunicação (Televisão, Rádio, Jornal, Internet, Mídias Sociais Digitais), as estratégias de rua, a comunicação direta feita por telefone e outros meios, bem como toda a parte de agendamento da campanha na imprensa, que era dirigida por uma equipe de jornalistas.

Do lado de fora do pólo existia também um amplo espaço externo, arborizado, com jardins, calçadas e amplos estacionamentos. O local era aproveitado também para a produção de programas eleitorais para a televisão com o candidato, seus familiares e seus apoiadores.

Ali também era oferecida uma estrutura para que os profissionais envolvidos na campanha pudessem realizar seus momentos de descanso e fazer suas refeições sem ter que sair do local do comitê¹²¹.

¹²⁰ Procurei registrar o lugar de uma forma que não identificasse as pessoas e nem tão pouco de qual candidato se trata a campanha. Por isso, a imagem não tem muito foco porque foi intencional.

¹²¹ Todos os dias por volta das 18 horas era oferecido um lanche para os profissionais envolvidos na campanha. Em todos os momentos em que estive fazendo a pesquisa no comitê, também participei desse lanche coletivo.

Segundo informações do gerente de marketing da campanha trabalhavam em torno de mil profissionais, entre jornalistas, cinegrafistas, editores, produtores de cinema, publicitários, analistas de mídias, analistas de dados, cientistas políticos, profissionais de marketing, cientistas sociais, profissionais da área de tecnologia de informação e profissionais de apoio (motoristas, serviços gerais, copeiras, secretárias, recepcionistas, seguranças). Entre esses profissionais estavam especialistas em campanhas eleitorais internacionais vindos de outros países (Portugal, Estados Unidos, Venezuela) e profissionais brasileiros com experiências em campanhas eleitorais internacionais e com formações em escolas internacionais como na University of Washington (UW)¹²². Entre as campanhas trabalhadas por esses profissionais estava a campanha americana de Barack Obama de 2012 e a campanha de Henrique Capriles da Venezuela na disputa eleitoral de 2013.

Ali ficavam reunidas em torno de oito empresas que eram gerenciadas pela Prole, principal agência de comunicação e marketing, comandada pelo publicitário Renato Pereira, que também era o gerente geral da campanha de comunicação do candidato. Além dessa, estava também a empresa Ideia Inteligência que tinha a frente o coordenador da equipe de inteligência de dados e informação, Maurício Moura¹²³, responsável pela coleta e análise de dados da campanha e que gerenciava todo o processo de *big data*, e a empresa Casa Digital, que cuidava de toda a campanha direcionada nas mídias digitais.

As equipes eram assim divididas:

- **Equipe de inteligência de dados:** composta por profissionais da área de pesquisa eleitoral, cientista político, cientista de dados, estatístico, gerentes e desenvolvedores da área de TI, jornalistas e profissionais de marketing.
- **Comunicação eleitoral:** cuidava do desenvolvimento de toda a campanha nas mídias tradicionais como televisão, rádio e impresso. Era composto por profissionais de marketing, publicidade e propaganda, jornalistas, fotógrafos, cinegrafistas.
- **Comunicação direta:** cuidava do desenvolvimento de mensagens e comunicações para públicos específicos da campanha, bem como fazia toda a campanha de rua.

¹²² A Universidade de Washington (em inglês, University of Washington / UW) é uma universidade pública de pesquisa em Seattle, Washington, Estados Unidos. Fundada em 1861, a UW é a maior universidade do noroeste americano e uma das mais antigas da costa oeste. É uma das universidades mais prestigiadas do mundo no que se refere ao ensino e à pesquisa.

¹²³ Maurício Moura é um dos diretores da empresa de dados Ideia Inteligência, em que fez a pesquisa na primeira fase de investigação do objeto.

- **Campanha digital:** cuidava do monitoramento e das ações de comunicação digital nos espaços da rede. Em média se dedicavam a esse trabalho em torno de 40 profissionais analistas de mídias sociais. Era considerada a maior equipe da campanha.

6.7.1 Inteligência de dados e informações eleitorais

A equipe de inteligência de dados da campanha comandada por Maurício Moura era composta por profissionais da área de ciência política, ciências sociais, marketing, jornalismo, economia, estatística, tecnologia da informação e direito. Cada profissional era responsável por uma atividade dentro da campanha do candidato, procurando integrar as suas ações com as ações dos demais profissionais. Parte dos profissionais ficavam instalados no comitê eleitoral, mas o restante como por exemplo da equipe de modelagem do sistema Alexandria, que era o sistema em que se operava o *big data*, ficava gerenciando tudo de dentro da própria empresa, que fica no centro do Rio de Janeiro. Como o sistema Alexandria era integrado e disponibilizado na nuvem do site da empresa, as equipes poderiam fazer o gerenciamento de dados e informações de outros lugares de forma virtual, sem ter que estar todos reunidos num mesmo espaço de trabalho.

Entre as atividades desenvolvidas pela equipe estava a de:

- (a) **Reunir dados de todos os tipos e fontes eleitorais;**
- (b) **Fazer pesquisas quali, quanti e de grupos focais;**
- (c) **Fazer combinações de informações de vários ambientes informacionais;**
- (d) **Traçar perfis individuais de eleitores por meio de técnicas e metodologias de análises de dados e *big data*;**
- (e) **Apontar os meios de comunicação para atingir os eleitores segmentados;**
- (f) **Direcionar as mensagens para os públicos de eleitores específicos.**

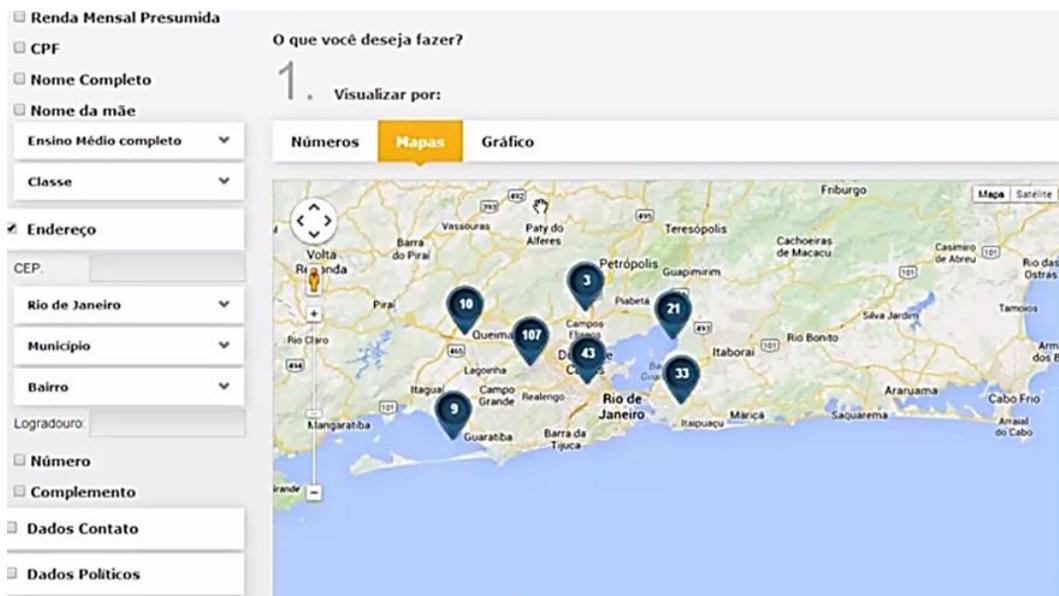
Para atingir esses objetivos, o analista de dados usou das técnicas do *microtargeting*, que tem por finalidade dividir o público/eleitores por nichos geográficos e sociais. Além disso, procurar informações que se aproxima da particularidade do comportamento eleitoral de cada cidadão. Depois da obtenção das coordenadas geográficas de uma determinada região do Rio de Janeiro, foi feita uma coleta de dados públicos. Com isso em mãos e com o auxílio de dados estatísticos – construíram-se diversas listas com detalhamento de todos os subgrupos desta região. Essas informações eram transferidas para um banco de dados do Sistema Alexandria que dá de forma clara e visualmente onde encontrar determinados eleitores, que podem ser analisados por segmentos e microsegmentos. Depois dessas visualizações por

nichos e variáveis era possível determinar e direcionar os tipos de ações que a campanha deveria seguir para atingir de forma satisfatória esses eleitores.

Os eleitores **inclinados a favor** e os **indefinidos** era os dois grupos onde a campanha concentrava os maiores esforços de comunicação. Porém, o “target”, ou eleitorado alvo, que é aquele que quantitativamente, somado ao contingente daqueles que já estão **crystalizados a favor**, não eram desconsiderados das estratégias de comunicação, porque produziam os votos necessários para eleger o candidato. Segundo Renato Pereira, o grau de detalhamento da campanha era tão grande que era capaz de determinar até mesmo a agenda pública do candidato. A campanha sabia por meio de estudos quais os bairros, cidades e regiões o candidato podia ir e o que falar e para quem falar, de acordo com os dados previamente levantados em pesquisas qualitativas pela equipe de inteligência de dados. Esse tipo de pesquisa buscava a segmentação dos eleitores. É importante lembrar que o objetivo da campanha eleitoral é de comunicar a mensagem certa para o eleitor certo. A segmentação do eleitorado, além de ser estratégia necessária, torna a campanha mais barata, já que o custo envolvido na conquista do voto varia dentro do eleitorado.

A Figura 18 ilustra a forma como a campanha trabalhava a partir dos dados coletados e analisados pelo sistema *big data*.

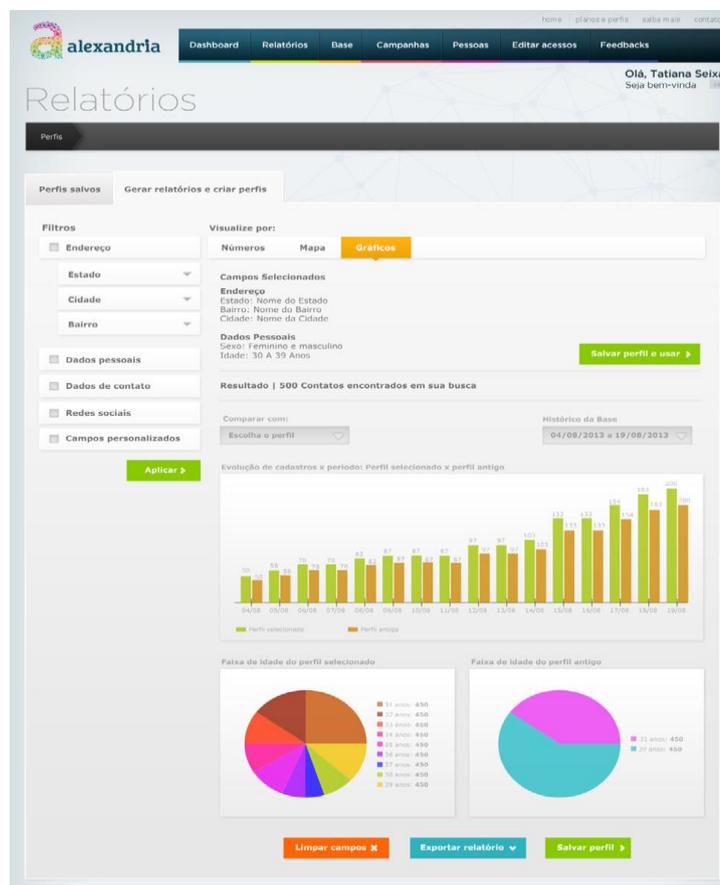
Figura 18: Exemplo do mapa geográfico dos eleitores



Fonte: Ideia Inteligência (2014)

Essa visualização poderia ser feita de várias formas, bastava uma filtragem por região, renda mensal, classe, grau de ensino. Enfim, como demonstrado nas variáveis na Figura 19.

Figura 19: Relatórios do Sistema Big data



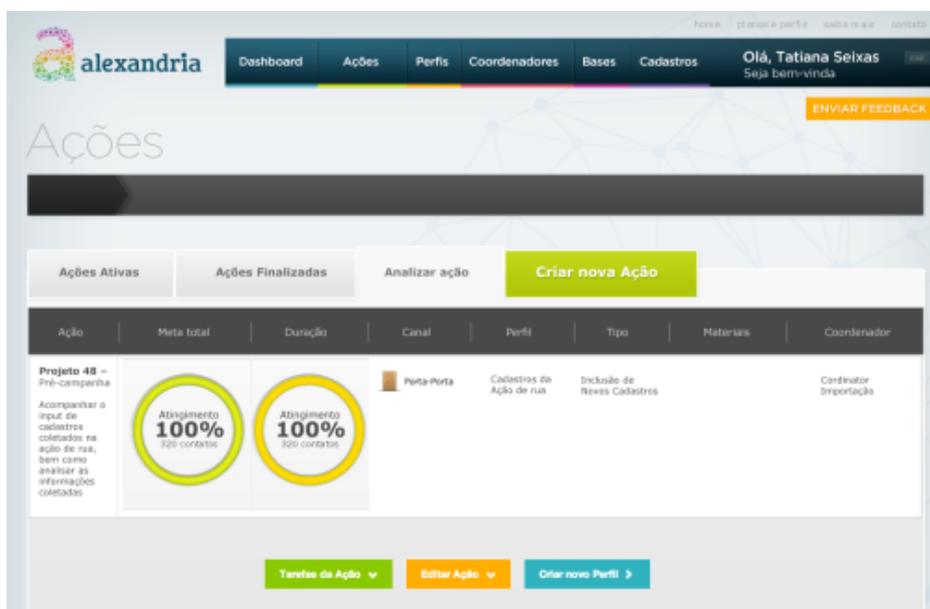
Fonte: Ideia Inteligencia (2014)

Outro tipo de ação que contribuiu para a coleta de dados e informações na campanha era o “porta a porta” os voluntários (cabos eleitorais) batiam na casa de eleitores em bairros, após um mapeamento inicial feito pela equipe de campanha. Segundo a coordenadora de pesquisas eleitorais na campanha, foram feitas visitas em aproximadamente 250 mil casas em todo o Estado, com o intuito de coletar dados e informações sobre os eleitores.

Para fazer as abordagens e pesquisas qualitativas com os eleitores foram contratadas pessoas com certas qualificações (a preferência era dada para cabos eleitorais com nível superior), pessoas com boa retórica, com poder de persuasão junto aos moradores de cada bairro do Estado. As ações eram feitas com o apoio de ferramentas e suportes tecnológicos como *tablets* com o propósito de agilizar o processo de coleta, captura, armazenamento e transmissão de dados e informações. Assim, na medida que as informações e dados eram coletados junto aos eleitores eram transmitidos instantaneamente para um banco de dados do

sistema central Alexandria. Além disso, o sistema fazia um gerenciamento sobre as ações realizadas nas ações dos coletores de dados, como mostra a Figura 20 .

Figura 20: Cadastro das ações de rua



Fonte: Ideia Inteligência (2014)

Além dessas ações de coleta de dados porta a porta, outras formas de coleta eram empreendidas na campanha. Segundo Maurício Moura¹²⁴ os dados que abasteciam a campanha, na sua grande maioria eram de domínio público.

“As bases de dados que usamos na campanha são dados de domínio público e de livre acesso. Buscamos respeitar a lei de acesso e uso de informações pessoais, sempre buscamos seguir o que diz a legislação do país” (MOURA).

Moura observa que no que se refere aos dados, à campanha foi trabalhada por etapas. A primeira etapa ocorreu durante a pré-campanha do **candidato X**, e se baseou em desenvolver pesquisas de campo e telefônicas que servissem de suporte para aumentar a base de dados já existente. Segundo o especialista essa etapa produziu aproximadamente uma base relevante de entrevistas considerando o tamanho do eleitorado local que era de 12 milhões de eleitores. No que se refere as pesquisas telefônicas foram feitas aproximadamente 150 mil entrevistas pelo sistema automático, que o eleitor respondia apenas apertando uma tecla. Com isso a equipe de dados atualizava a base de dados sobre a população eleitoral do Estado, já que nas campanhas de 2010 (eleições para governo) e 2012 (eleições para prefeito) foi feito um esforço para desenvolver uma base de dados para essas campanhas e os dados poderiam

¹²⁴ Maurício Moura fez declarações à pesquisadora durante os momentos de observação da rotina da campanha do candidato.

ser (re)aproveitados. Isso de certa forma pode ser denominado de datificação (SCHÖNBERGER-MAYER; CUKIER, 2013), porque além de coletar os dados e informações para determinada finalidade (uso primário), anos depois esses dados passaram por uma atualização e foram reaproveitados para outra campanha (usos secundários).

Além dessa coleta, profissionais especialistas em pesquisas eleitorais realizaram entrevistas com eleitores de todo o estado, que incluíam perguntas eleitorais como: **Se a eleição fosse hoje em que o senhor (a) votaria? E em quem o senhor(a) definitivamente não votaria?**

A pesquisa também procurava dimensionar e qualificar o governo, englobando perguntas sobre aprovação/avaliação de governo. Ex: **E como senhor(a) avalia o atual governo? A escolha das opções de respostas ficava entre: ótimo, bom, regular, ruim e péssimo.**

Para um dos profissionais essa coleta de dados é considerada essencial numa campanha eleitoral.

“Ao recolher o sentimento dos cidadãos em relação ao governo, a equipe responsável pelas estratégias de comunicação já pode direcionar algumas ações de marketing da campanha e trabalhar estrategicamente e de forma segmentada com eleitores satisfeitos ou insatisfeitos com a administração do atual governo” (CS).

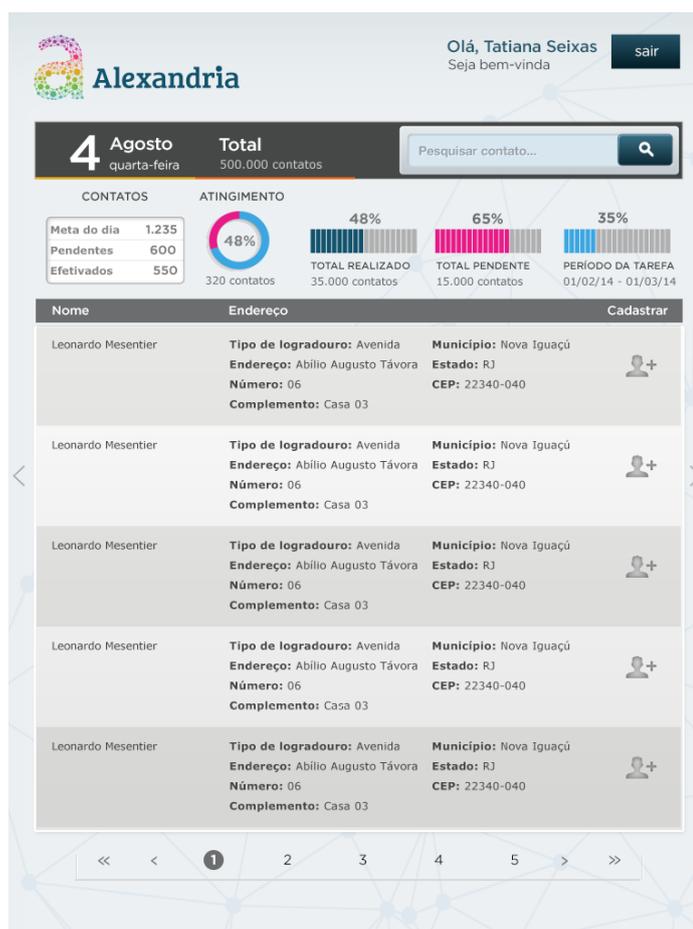
Todas as informações obtidas por meio de pesquisas eram jogadas em um formulário digital (Figura 21) que após preenchido na pesquisa de campo era transmitido em tempo real para um banco de dados do Sistema Alexandria, que posteriormente eram utilizados para as análises conjuntas com os demais dados da campanha.

Figura 21: Página de cadastro sistema *big data*

The image shows a web interface for the 'Alexandria' system. At the top, there's a header with the logo and the name 'Alexandria'. Below that, a navigation bar shows the date '4 Agosto quinta-feira' and a 'Total' of '500,000 contatos'. A search bar is present with the text 'Pesquisar contato...'. The main content area is titled 'Cadastrar' and contains a registration form. The form fields are filled with the following information: Nome: Leonardo Meserlier; CEP: 22460-040; Tipo de Logradouro: Avenida; Endereço: Abílio Augusto Távora; Número: 06; Complemento: Casa 03; Município: Nova Iguaçu; Estado: RJ; Data de nascimento: (empty); Renda Mensal: (empty); Profissão: (empty). There is a profile picture placeholder with the text 'Incluir foto'. At the bottom of the form, there are radio buttons for 'Aprova fortemente', 'Aprova', 'Desaprova', and 'Não sabe'. The 'Aprova fortemente' option is selected. Below the radio buttons are two buttons: 'Limpar' with a red 'X' icon and 'Salvar' with a green arrow icon.

Fonte: Ideia Inteligência (2014)

Figura 22: Exemplo de cadastro do sistema



Fonte: Ideia Inteligência

A Figura 22 ilustra o exemplo de cadastro que era utilizado para armazenar dados dos eleitores do Estado do Rio de Janeiro. O sistema Alexandria tinha por objetivo organizar os dados cadastrados a partir de municípios. Além disso, ficavam registrados dados como nome do eleitor e endereço para futuras ações de comunicação eleitoral.

O profissional responsável pela modelagem do Alexandria diz que no sistema é possível armazenar outros tipos de dados dos eleitores. As informações detalhadas dos eleitores servem para a campanha como se fosse um “DNA” do cidadão. Para compor a base de dados da campanha eram capturadas as seguintes informações do eleitor: **nome completo, endereço completo, telefone(s) fixo e celular, idade, sexo, nome da mãe, renda mensal presumida, zona e seção eleitoral, situação do título, e até mesmo a inclusão da foto do eleitor cadastrado.**

“Cada informação pessoal passa a ser considerada uma variável a ser trabalhada no conjunto de dados de eleitores” (MOURA)

O objetivo principal nesse tipo de pesquisa é mensurar e qualificar as principais variáveis que mais afetavam um eleitor na sua particular decisão de votar em um determinado candidato (**candidato X**) ou seus principais adversários.

Para chegar a essas respostas, Moura usa uma técnica que tem embasamentos nos modelos econométricos de *Logíst* e *Probit*, que é especificado pela seguinte fórmula (Figura 23). Esses modelos são usados em ciências do comportamento e desenvolvida por Luce e Suppes e traduzida econometricamente por MaFaden (ALDRICH e NELSON, 1984).

Figura 23: **Fórmula de probabilidade**

$$y^* = \sum x\beta + e \quad e|X \sim Normal(1, 0)$$

Fonte: Moura (2011)

A fórmula é explicada por Moura num artigo publicado em 2011, da seguinte forma: Onde y^* = vota ou não em um determinado candidato, X = determinada resposta a cada pergunta nas pesquisas – sexo, idade, aprovação, avaliação e etc, e B = coeficiente de cada variável afetando a decisão do voto, ou seja, quanto maior mais essa variável influencia na decisão do voto.

Assim com as variáveis-chaves de intenção de votos determinadas (não divulgadas na pesquisa por determinações da gerencia da campanha e do partido político) a equipe de inteligência de dados criou uma classificação de probabilidade em votar no **candidato X** para cada entrevistado nas pesquisas eleitorais. Simplificando essa explicação do especialista, seria que a cada pesquisa/tracking eleitoral¹²⁵ realizada pela estratégia de campanha, em função das respostas de um determinado eleitor a diversas perguntas, era possível estimar a probabilidade entre 0 e 1¹²⁶ desse eleitor votar no **candidato X** e também nos adversários. O tracking é uma modalidade de pesquisa que busca monitorar a volatilidade da opinião pública com base em pesquisas feitas diariamente por telefone durante todo o processo eleitoral. Essas estimativas foram amplamente utilizadas e monitoradas durante a campanha. Esse monitoramento se dava

¹²⁵ A empresa que fazia a pesquisa tracking não foi identificada e revelada na pesquisa por questões éticas e de anonimato .

¹²⁶ Usa-se para isso a lógica booleana que foi originalmente desenvolvida pelo matemático inglês George Boole em meados de 1838. A lógica permite que uma série de coisas pouco prováveis sejam mapeadas em bits e bytes.

por diversas formas: pesquisas quantitativas, pesquisas qualitativas, pesquisas com grupos focais, pesquisas de intenção de votos pelos institutos oficiais, monitoramento de mídias sociais.

Um dos exemplos citados para entender como se dá a probabilidade do voto seria de que um indivíduo que aprova e avalia bem um governo pleiteando reeleição tem maior probabilidade de votar em um candidato de situação. Para o especialista essa constatação é puramente matemática, porém não tão simples assim de se entender.

“Essa modelagem possibilita mensurar matematicamente o quanto determinado eleitor é mais propenso a votar e quais são suas preferências, anseios e características, o que facilita imensamente o trabalho da equipe de comunicação” (MOURA).

Diante da modelagem foi agrupado e monitorado os eleitores em função da sua probabilidade de votar no **candidato X**. A probabilidade era distribuída entre 0 e 1. Assim a equipe de inteligência de dados da campanha tinha de forma resumida e simplificada três tipos de grupos:

- (1) Grupo com altíssima probabilidade de voto;**
- (2) Grupo de probabilidade intermediária;**
- (3) Grupo de baixa probabilidade.**

Segundo Maurício Moura para todos os grupos era possível saber exatamente as características principais deles e quais eram as variáveis que os afetam mais a aprovação de governo para eles, como segurança, saúde, educação, viabilidade urbana, entre outros.

O importante dessa metodologia é que ele ajuda na construção da mensagem e na mensagem do plano de mídia da campanha e que permite a campanha falar diretamente com esses grupos pré-definidos antes mesmo da campanha estar na rua (MOURA).

Segundo Moura usar esse tipo de metodologia na campanha permitiu uma leitura mais detalhada do eleitorado do Estado do Rio de Janeiro e a equipe de campanha passou a ter uma ferramenta estratégica para guiar as suas ações eleitorais. Além disso, pode-se ter um impacto direto na comunicação desenvolvida para o **candidato X** de diversas formas:

- (a) Na otimização do conteúdo audiovisual;**
- (b) Melhor aproveitamento do plano de mídia;**
- (c) Maior foco no trabalho de campo;**
- (d) Um aproveitamento mais efetivo do candidato.**

Na avaliação de Moura isso tudo resultou em um direcionamento mais produtivo de esforços e energia para atingir de maneira efetiva cada grupo de eleitores em que a pesquisa apontou no processo inicial e que foi se mantendo atualizado ao longo da campanha eleitoral,

com a inclusão de novos dados coletados que vinham de outros ambientes de forma instântanea como da internet, principalmente das redes sociais digitais.

6.7.2 De onde vêm os dados da campanha

Em relação aos dados, não foi possível ter acesso aos bancos de dados diretamente, por ser considerados dados sigilosos e pertencentes ao cliente (**candidato X**). Essa era a recomendação principal da gerência de inteligência de dados. No entanto, pelo acompanhamento a partir das observações empreendidas com alguns profissionais de dados e a partir das entrevistas realizadas durante a permanência da pesquisadora no comitê, foi possível entender de onde estavam vindo os dados e informações que abasteciam as estratégias da campanha.

Segundo os profissionais envolvidos com a coleta e armazenamento, os dados eleitorais vinham dos bancos de campanhas passadas, de institutos de pesquisa como IBGE (Instituto Brasileiro de Geografia e Estatística), do banco de dados do TRE (Tribunal Regional Eleitoral) ou do TSE (Tribunal Superior Eleitoral), do partido do candidato, do governo estadual e municipal, do banco particular da empresa de inteligência, das pesquisas internas da campanha como: pesquisas feitas por telefone; pesquisas de rua que estavam sendo realizadas durante a campanha; dos grupos focais que eram administrados pela própria campanha e das mídias sociais digitais, por meio de monitoramentos de redes sociais, que foram realizados na pré-campanha e durante a campanha eleitoral.

Todas as ações de coleta de informações eram remetidas a uma base de dados diretamente no banco de dados do sistema Alexandria.

O profissional que cuidava da campanha de rua do candidato (CR), explicou que a campanha também se beneficiou de um projeto que tinha por objetivo funcionar como um “gabinete itinerante”. Esse projeto era uma espécie de governo avançado, instalado em diversas comunidades do Estado do Rio de Janeiro. O objetivo era atender às demandas da população a respeito de qualquer situação da administração pública. A partir de cadastros que eram feitos com moradores das comunidades, em que ficavam registrados seus dados pessoais, era possível identificar os tipos de problemas de determinada localidade e esses temas e informações eram utilizadas posteriormente durante a campanha do **candidato X**.

O candidato juntamente com os assessores que fazia o atendimento à população, ouvindo e interagindo com as pessoas sobre os seus problemas e necessidades de governo.

“Além do atendimento às necessidades dos cidadãos cariocas essa era uma forma de se captar informações que serviram posteriormente para as estratégias de campanha durante as eleições de 2014” (CCR).

A respeito das informações e dados coletados durante a campanha eleitoral, Maurício Moura ressalta que quanto mais variáveis a campanha tiver sobre o perfil dos eleitores, maior é o conhecimento sobre o segmento em que o político deverá ser inserido.

O especialista também destaca que para compor os bancos de dados eleitorais se utilizou das listas amarelas (antigas listas telefônicas impressas). Essas listas eram utilizadas para coletar dados como nomes, endereços, telefones e tipos de atividades que os eleitores estão envolvidos, se são profissionais liberais, comerciantes, entre outros. São listas consideradas públicas, porque os usuários da telefonia fixa autorizam a divulgação pública de alguns de seus dados pessoais e profissionais. Essas listas servem para ações de comunicação segmentada e direta (*one to one*), como envio de malas diretas, chamadas automática telefônicas promovendo a campanha e também para pesquisas *tracking*.

6.7.3 Estrutura do trabalho de *big data*

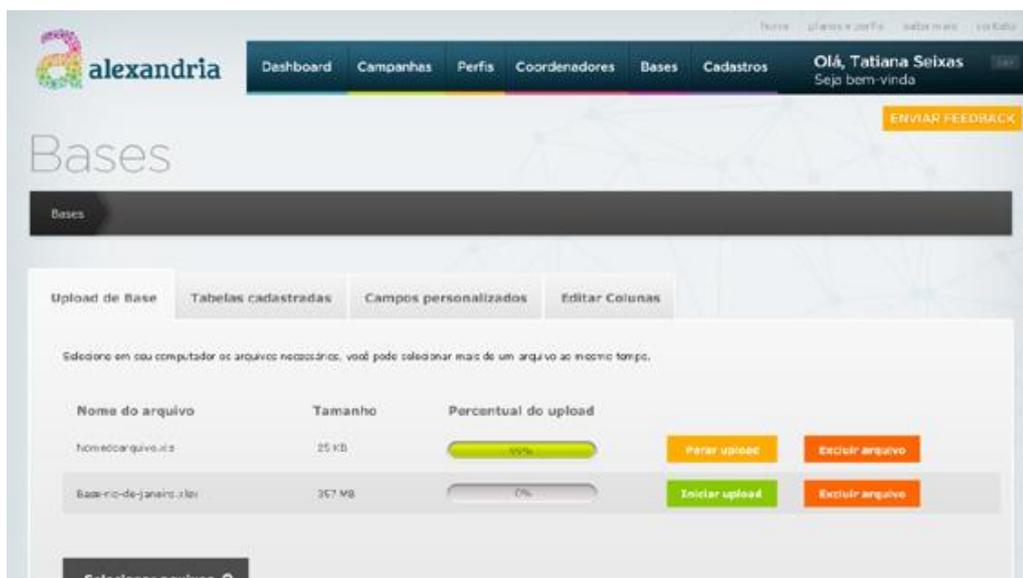
O trabalho de *big data* é contínuo desde a coleta de dados até a sua análise. Os dados são realimentados de fontes *off line* e *on line*. A preocupação da equipe de inteligência era entender o que os eleitores desejavam dos seus candidatos, para isso eram analisados todos os ambientes de informação. Com esses dados pode-se tirar o maior número de *insights* para a campanha do candidato.

Na campanha eleitoral, como já citado foi usado o sistema Alexandria, que atende a todas as fases do ciclo de vida do *big data*, que vai desde a coleta, criação, higienização e alimentação de uma base de dados até a análise e cruzamento. Segundo o especialista em Tecnologia da Informação e também responsável pelo sistema Alexandria durante a campanha eleitoral, essa base não têm limites de tamanho ou número de campos, que podem ser customizados de acordo com a necessidade de cada campanha eleitoral.

“O Alexandria é um sistema que te permite trabalhar de maneira dinâmica e integrada, automatizando todos os processos relacionados a construção, gerenciamento, alimentação e higienização da sua base de dados, integrando informações de diferentes fontes e formatos em um único banco” (ETI).

No exemplo que segue da Figura 24 contempla a característica da **variedade** do *big data*, porque o sistema usado na campanha eleitoral era capaz de integrar diferentes tipos de dados, estruturados, semi-estruturados e não estruturados e a partir daí compor diversos *data sets*. Todo esse processo passa por uma limpeza dos dados feitas por um profissional da área de estatística.

Figura 24: Exemplo de uma base de dados



Fonte: Ideia Inteligência (2014)

Nesse módulo ilustrado pela Figura 24 é possível fazer os cruzamentos de informações de diferentes planilhas através de um ou mais campos em comum. Por exemplo, é possível cruzar planilhas de monitoramento on-line (redes sociais) com base de contatos off-line de eleitores. A Figura 25 mostra como é possível moderar os dados cadastrados.

Figura 25: Moderação do big data

Data de inserção	Operador	Tarefa	Conflito
DD/MM/AAAA	Felipe Ferreira Vieira	T1: Nome da Tarefa a ser realizada	3 contatos
DD/MM/AAAA	Felipe Ferreira Vieira	T1: Nome da Tarefa a ser realizada	3 contatos
DD/MM/AAAA	Felipe Ferreira Vieira	T1: Nome da Tarefa a ser realizada	3 contatos
DD/MM/AAAA	Felipe Ferreira Vieira	T1: Nome da Tarefa a ser realizada	3 contatos
DD/MM/AAAA	Felipe Ferreira Vieira	T1: Nome da Tarefa a ser realizada	3 contatos
DD/MM/AAAA	Felipe Ferreira Vieira	T1: Nome da Tarefa a ser realizada	3 contatos
DD/MM/AAAA	Felipe Ferreira Vieira	T1: Nome da Tarefa a ser realizada	3 contatos
DD/MM/AAAA	Felipe Ferreira Vieira	T1: Nome da Tarefa a ser realizada	3 contatos
DD/MM/AAAA	Felipe Ferreira Vieira	T1: Nome da Tarefa a ser realizada	3 contatos
DD/MM/AAAA	Felipe Ferreira Vieira	T1: Nome da Tarefa a ser realizada	3 contatos
DD/MM/AAAA	Felipe Ferreira Vieira	T1: Nome da Tarefa a ser realizada	3 contatos
DD/MM/AAAA	Felipe Ferreira Vieira	T1: Nome da Tarefa a ser realizada	3 contatos
DD/MM/AAAA	Felipe Ferreira Vieira	T1: Nome da Tarefa a ser realizada	3 contatos
DD/MM/AAAA	Felipe Ferreira Vieira	T1: Nome da Tarefa a ser realizada	3 contatos
DD/MM/AAAA	Felipe Ferreira Vieira	T1: Nome da Tarefa a ser realizada	3 contatos

Fonte: Ideia Inteligência (2014)

Dessa forma, novos dados inseridos são adicionados à base existente e, caso haja algum tipo de conflito, o usuário pode mesclar cadastros através do módulo Moderar (Figura 25), garantindo a unidade de informações do seu banco de dados.

Outra característica importante do *big data* é de que o sistema integra o **grande volume** de dados. Em vez de se trabalhar com base em amostragens de eleitores, como eram feitas as campanhas políticas anteriores. O sistema tem a capacidade de comportar a base completa de eleitores do estado do Rio de Janeiro (em torno de 12 milhões de eleitores em 2014), divididos por nichos estabelecidos previamente na campanha, conforme os objetivos das conquistas de votos.

“O Alexandria tem por objetivo fundir as múltiplas identidades dos cidadãos fluminenses, os militantes do partido envolvidos na campanha do candidato, o eleitor off-line e on-line, o voluntário em um único perfil político unificado. Os dados armazenados num mesmo local é capaz de estabelecer comunicação uns com os outros” (TI).

Segundo a profissional de marketing da empresa responsável pelo Sistema Alexandria, esse é o grande diferencial do *software*. Integrar diferentes bases em um único banco e conseguir visualizar estas informações de diferentes maneiras para basear planejamentos estratégicos na busca do voto.

“Há uma série de filtros possíveis e que são customizados de acordo com o interesse de cada candidatura” (GM).

6.7.4 As mídias sociais na campanha

A internet não é apenas um canal de comunicação entre candidatos e eleitores, o grande benefício é o poder de coleta de informações, segmentação dos destinatários da mensagem, a mensuração imediata do resultado da ação por meio dos posts, *Twitter*, *Facebook* etc (CAPELLA 2014). E isso foi exatamente o cenário da campanha eleitoral em que foram feitas as observações de campo.

Todas as relações de comunicação com os eleitores eram aproveitadas pela equipe de comunicação digital, não só para estabelecer relações interativas entre o candidato e os eleitores, mas também na captação de informações e direcionamento das mensagens.

Na opinião de um dos profissionais envolvidos na campanha digital, as mídias digitais se tornaram um elemento vital na conquista do eleitorado campanha de 2014. Segundo o especialista os objetivos da campanha digital eram:

- (a) Captar dados de eleitores para ações na campanha;**
- (b) Conseguir um determinado número de seguidores em sites de mídias sociais que poderiam se tornar futuros eleitores;**

- (c) Engajar pessoas e criar uma rede de voluntários para colaborar na campanha;**
- (d) Divulgar ideias e testar a atratividade dos temas que o candidato estava propondo;**
- (e) Observar tendências e coletar ideias para a campanha;**
- (f) Interagir com o público e criar espaço para o debate de suas ideias.**

Para alcançar todos esses objetivos, os profissionais trabalhavam dia e noite coletando e analisando dados da campanha nas mídias sociais, por meio de ações de monitoramento, utilizando *software* específicos para isso.

“O monitoramento de mídias sociais é um termômetro do que as pessoas estão falando sobre o candidato ou assunto e traz informações de um jeito mais rápido e instantâneo para a campanha do que as pesquisas eleitorais tradicionais” (AMD, 2014).

Durante a pesquisa no comitê eleitoral foi possível observar algumas dessas ações. Os profissionais ficavam atentos ao fluxo de informações que circulavam pelas redes sociais sobre a campanha. Além disso, também era feito um monitoramento qualitativo nas páginas oficiais do candidato. Segundo um dos profissionais responsáveis por essas ações, o monitoramento das redes sociais têm propósitos claros para toda a equipe.

“Os programas na televisão exigem uma atenção especial. É preciso fazer a monitoria do que está sendo falado em tempo real. Na campanha, ao final de cada programa, soltamos um boletim com o volume de conversações geradas nas mídias sociais, a análise de sentimento para aquele período e os itens mais representativos. O trabalho de tagueamento dos dados ocorre posteriormente” (AMD).

Taguear vem da palavra TAG em Inglês quer dizer etiqueta. Utilizada na web visa ordenar, classificar, organizar informações, através de palavras chave, dentro de uma hierarquia de informações que permite relacionar significados semelhantes. Estas palavras chaves são conhecidas como Tags ou metadados, cuja finalidade é classificar e organizar arquivos, páginas e conteúdos, visando encontrar, com facilidade, aquilo que o usuário deseja.

Ainda no que se refere ao monitoramento das mídias sociais o especialista ainda completa a sua fala dizendo:

“Em dias de debate, o processo é um pouco diferente do que em outros dias. O trabalho mais intenso nas redes sociais. Como o volume de conversas numa situação dessas é bem maior, quase toda a equipe trabalha no acompanhamento do monitoramento. Nossos analistas transmitem para os assessores do candidato o que está acontecendo, indicando quais falas estão sendo bem recebidas ou não. De certa forma, as informações trazidas pelo monitoramento são um termômetro muito melhor que os grupos focais, selecionados pela consultoria de marketing para saber qual candidato tem um desempenho melhor no debate” (GCD, 2014).

O monitoramento é uma tecnologia de pesquisa acessível, que permite conhecer a opinião espontânea e em tempo real, de milhares de pessoas, sobre qualquer assunto e sobre os candidatos. O profissional ainda explica porque é importante fazer monitoramento dos candidatos adversários.

“É fundamental monitorar a concorrência. Fazemos o tagueamento das informações referentes aos adversários do candidato, bem como a análise de sentimento. Também realizamos alguns trabalhos pontuais, como mapear

os ativistas mais influentes nas mídias sociais da campanha adversária, ou seja, pessoas cujas mensagens e ideias eram replicadas – independentemente de eles terem 100 ou 1000 seguidores. Identificar essas pessoas não é um trabalho simples. Você vai descobrindo quem elas são aos poucos, ao longo da campanha. É uma tarefa que exige uma análise específica das informações do monitoramento. A partir do momento em que você sabe quem são os ativistas, pode ficar focado no que escrevem, produzindo conteúdos específicos para responder a eles” (GCD).

Assim as ações nas mídias sociais eram delineadas em três estágios diferentes: **Visibilidade, Interação e Conversão**. O funil (Figura 26) ajuda a entender quais eram as metas pretendidas nas ações on-line da campanha do candidato analisado.

Figura 26: Funil da participação do eleitor nas redes sociais



Fonte: Adaptado de manuais de monitoramento de mídias sociais

A **visibilidade** diz respeito à relação passiva com o eleitor; a **interação** é a relação interativa com eleitor; e a **conversão** é o eleitor integrado à campanha (RECUERO, 2012).

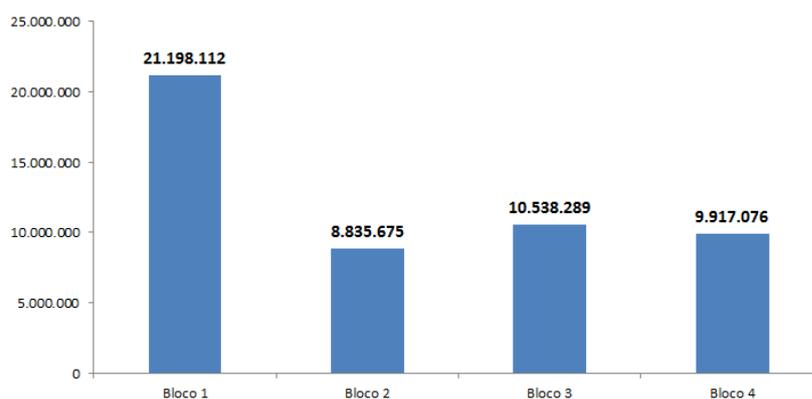
Um dos trabalhos observados durante a pesquisa no comitê foram às ações da equipe de comunicação digital a partir do acompanhamento do debate político que acontecia na TV e a repercussão nas redes sociais. Todo o trabalho era feito a partir de dados que estavam disponíveis na internet, ou seja, de manifestações que as pessoas produziam independentemente de estímulos ou direcionamentos. Muitas das menções nas redes sociais se dava instantaneamente a partir de temas que eram debatidos na televisão pelos candidatos.

O trabalho era realizado de forma integrada e em tempo real. Ficava reunido num mesmo ambiente um total de 10 profissionais entre analistas de mídias sociais, profissionais de jornalismo, marketing e designer, que eram responsáveis por abastecer as redes de comunicação do candidato com informações pertinentes aos temas que estavam sendo debatidos no programa de televisão ao vivo.

Além disso, uma equipe de profissionais ficava responsável por responder aos *posts* e mensagens de seguidores do candidato em suas páginas. Nenhum *post* ou nenhuma mensagem ficava sem respostas durante o período em que o debate político estava acontecendo. A cada bloco do debate, a equipe responsável pelo monitoramento das mídias

sociais gerava um relatório, informando ao coordenador, qual era o desempenho do candidato e dos adversários, bem como as repercussões que o debate vinha tendo nas redes. O gráfico representado pela Figura 27 representa um dos resultados desses monitoramentos¹²⁷.

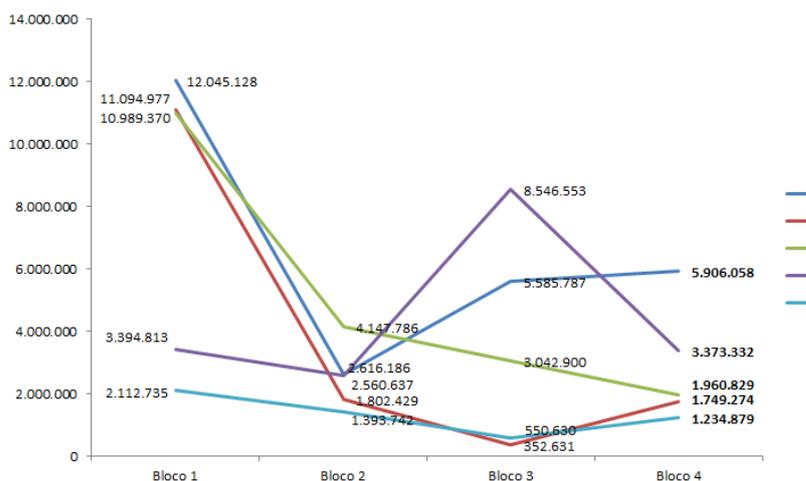
Figura 27: Alcance do programa de debate eleições para governo do estado do Rio de Janeiro 2014



Fonte: Campanha do candidato

Neste tipo de monitoramento de mídias sociais o objetivo era entender como se desenvolvia o debate na TV e qual era o alcance que o programa estava tendo nas redes sociais. O monitoramento era feito bloco a bloco e ao final do debate era gerado um resultado final com todos os blocos do debate político como mostra o Gráfico ilustrado pela Figura 28

Figura 28: Desempenho dos candidatos em cada bloco do debate



Fonte: Material da Campanha Candidato X (2014)

¹²⁷ O material foi liberado por um dos chefes da equipe de comunicação digital para que pudesse ser usado como ilustração na pesquisa deste trabalho.

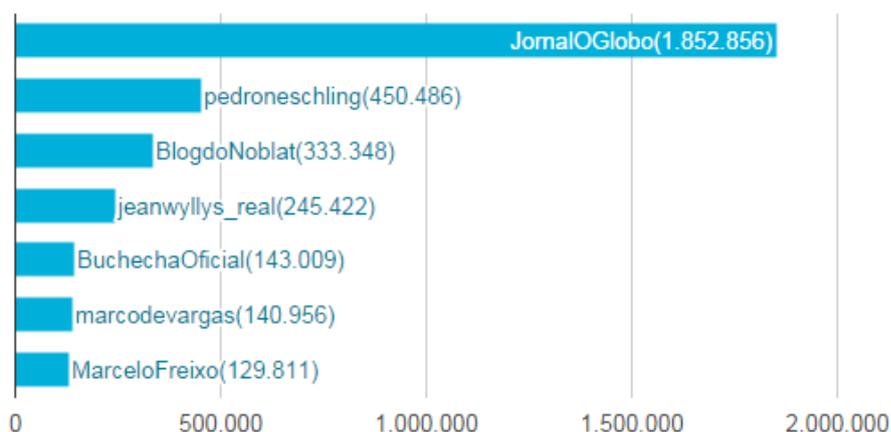
No gráfico (Figura 28) a cor azul escuro representa o **candidato X**, as demais cores são representadas por cada um dos candidatos concorrentes monitorados.

Assim, o monitoramento não dizia respeito apenas ao alcance do **candidato X**, mas também o alcance que os outros candidatos concorrentes estavam tendo nas redes sociais no momento em que o debate estava acontecendo na emissora de TV. As informações e dados coletados nesse monitoramento era feito somente sobre os candidatos que apresentavam nas pesquisas eleitorais com mais de 1% de intenção de votos.

O monitoramento buscava saber quem eram os influenciadores nas redes. As informações obtidas tinham por objetivo identificar na massa de dados e informações das redes quem eram os ativistas, simpatizantes e detratores do candidato.

No Gráfico representado pela Figura 29 que segue o Jornal O Globo se configurava como um dos maiores influenciadores com 1.852.856 menções, seguido por Pedro Neschling (450.486), ator e filho da atriz Lucélia Santos.

Figura 29: Os influenciadores da campanha



Fonte: Campanha do Candidato X (2014)

A nuvem de palavras (Figura 30) era sempre usada nos relatórios para identificar dentro de todo o monitoramento o destaque dos assuntos por palavras mais usadas.

Figura 30: Nuvem de palavras



Fonte: Campanha do Candidato

Nuvem de palavras também é chamada de nuvem de tags e consiste numa técnica que permite observar visualmente quais palavras foram mais citadas no conteúdo analisado no monitoramento. Para Manovich (2010, p. 12) a nuvem de *tags* mostra as palavras mais comuns em um texto no tamanho de letra correspondente à sua frequência no texto.

Na nuvem usada (Figura 30) como exemplo é possível visualizar que o debate político na televisão e repercutido nas mídias sociais teve mais menções sobre as palavras “**rabo**”, “**Tarcísio**”, “**Amarildo**”, entre outros. Nesse debate o candidato Tarcísio Mota do PSOL, direciona uma pergunta para o candidato que concorria à reeleição para governo Luiz Fernando Pezão (PMDB) sobre onde estava Amarildo.¹²⁸ No monitoramento das mídias sociais daquele dia foi considerado o momento mais comentado e mencionado.

No artigo *What is Visualization*, Lev Manovich (2010) diz que a visualização de informações está se tornando mais do que um conjunto de ferramentas, tecnologias e técnicas para grandes conjuntos de dados. Manovich explica que visualização de dados, historicamente, sempre envolveu a redução. Ao invés de mostrarmos uma lista com cinco mil respostas a uma *survey*, transformamos essa lista em um gráfico como um histograma, mostrando a distribuição das respostas, ou seja, uma redução. Esta redução envolve

¹²⁸ Amarildo Dias de Souza é um ajudante de pedreiro brasileiro que ficou conhecido nacionalmente por conta de seu desaparecimento, desde o dia 14 de julho de 2013, após ter sido detido por policiais militares e conduzido da porta de sua casa, na Favela da Rocinha no Rio de Janeiro, em direção a sede da Unidade de Polícia Pacificadora do bairro. Seu desaparecimento tornou-se símbolo de casos de abuso de autoridade e violência policial e motivos de muitos movimentos de protestos na cidade. Os principais suspeitos no desaparecimento de Amarildo são da própria polícia. Entre os dias 13 e 14 de julho de 2013, uma operação batizada de Paz Armada mobilizou 300 policiais na Rocinha e prendeu suspeitos sem passagem pela polícia, logo depois de um arrastão ocorrido nas proximidades da favela, e de acordo com a polícia, 30 pessoas foram presas, entre elas Amarildo. Ele havia acabado de voltar de uma pescaria e foi detido e conduzido por policiais militares da Unidade de Polícia Pacificadora (UPP) da Rocinha na noite do dia 14. Liberado após poucos minutos na delegacia após concluída a averiguação na UPP da Rocinha, desde então não se conhece o paradeiro do pedreiro. Dois dias depois, a família registrou o seu desaparecimento.

transformar um dado em outro formato visual: o volume de respostas a cada pergunta, por exemplo, se é traduzido/reduzido em tamanho de colunas.

Assim, usando dessas visualizações, no monitoramento feito na campanha, ao final eram destacados alguns *posts* e comentários das redes para exemplificar o que os usuários estariam falando sobre o debate daquele dia, tanto dentro de uma análise de sentimento positiva ou negativa em relação ao candidato e aos adversários (Figura 31).

Figura 31: Opinião nas redes

The figure displays three screenshots of tweets from Twitter, illustrating public opinion during a political campaign. Each tweet includes the user's profile picture, name, handle, and a 'Seguir' (Follow) button. The tweets are as follows:

- Tarcísio Motta (@MottaTarcisio):** "Sou eu que tenho rabo de cavalo, mas é ele (Pezão) que tem rabo preso com as empresas milionárias. #TarcisioNaGlobo". This tweet has 199 retweets and 100 likes, dated 00:11 - 1 de out de 2014.
- Luiz Antonio Alves (@Lasalves):** "#FelipePezao @LFPezao O Pezão está mostrando-a competência de sempre. O Tarcisio está desesperado. Toim na cabeça dele de novo." This tweet has 8 retweets and 4 likes, dated 00:10 - 1 de out de 2014.
- Marcela Carvalho (@marcela_penna):** "Pezão segue a lei!!! Se quiser saber mais, vamos procurar na internet!!! Que mala este Tarcísio!! #DebateNaGlobo". This tweet has 10 retweets and 2 likes, dated 00:12 - 1 de out de 2014.

Fonte: Equipe de Campanha

A equipe de monitoramento de redes sociais, também realizava relatórios gerais sobre o desempenho do **candidato X** e o cenário da campanha eleitoral, apresentando os desempenhos dessas análises em todos os debates políticos realizados. Esses relatórios indicavam que durante os debates foram feitas 56.036 menções sobre a campanha eleitoral de 2014. Essas menções vinham de usuários de Twitter (90%) e dos usuários do Facebook

(10%). Um ponto importante nessa constatação diz respeito ao grande uso do Twitter pelos eleitores para manifestarem suas opiniões sobre o debate. O Twitter nesse caso é usado tanto para divulgar a campanha quanto para coletar informações e sentimentos dos eleitores, que acabam contribuindo para os (re)direcionamentos das ações do marketing eleitoral de forma integrada.

O Quadro 16 mostra os temas e menções mais citadas nas redes sociais durante os debates do primeiro turno.

Quadro 16 : Temas abordados nas redes nos debates do primeiro turno eleições 2014

TEMAS	MENÇÕES
Segurança	3.884
Educação	1505
Saúde	1204
Corrupção	1091
Mobilidade Urbana	828
Administração Pública	738
Infraestrutura	698
Trabalho	267
Meio Ambiente	190
Habitação	159
Social	50
Transporte	49
Esporte	46

Fonte: Elaboração da autora a partir de dados da campanha

No relatório apresentado sobre os temas mais discutidos aponta que **segurança, educação, saúde e corrupção** eram as principais menções feitas pelos eleitores em relação as propostas apontadas pelos candidatos nos debates. Outro relatório que foi gerado sobre o desempenho do candidato durante os quatro debates políticos realizados no primeiro turno da campanha eleitoral de 2014, fazia referência às ocorrências e alcance que o evento foi comentado nas redes sociais. O Quadro 17 mostra como foi esse resultado.

Quadro 17: Comparativo – Ocorrências X Alcance no primeiro turno

Emissora	Dia	Ocorrências	Alcance
BAND		14.863	28.019.706
REDE TV		4.366	27.217.969
RECORD		5.426	41.305.113
GLOBO		31.381	102.996.463

Fonte: adaptado pela autora a partir de informações da campanha

Sobre o alcance dos debates o Quadro 17 mostra que o primeiro debate que foi realizado pela Band Rio no dia 19 de agosto de 2014 e o último debate realizado pela Rede Globo no dia 30 de setembro foram os que tiveram maior alcance nas redes sociais. O debate promovido pela Band Rio foi o primeiro debate a contar com a participação dos usuários Twitter, os eleitores ainda puderam participar através de perguntas enviadas pelo aplicativo WhatsApp. Como demonstra o Quadro 18 o segundo debate ocorreu no dia 02 de setembro e foi promovido pela RedeTV!, Jornal O Dia e portal IG. Esse debate foi realizado no Teatro Oi Casa Grande, no bairro do Leblon. Já o terceiro debate, promovido pela Record, foi realizado nos estúdios da emissora em 26 de setembro. E o último debate realizado pela Rede Globo aconteceu nos estúdios da emissora.

Quadro 18: Os debates no primeiro turno para governo do Rio de Janeiro

Data	Organizadores	Mediador	Anthony Garotinho (PR)	Luiz Fernando Pezão (PMDB)	Marcelo Crivella (PRB)	Lindberg Farias (PT)	Tarcísio Motta (PSOL)	Dayse Oliveira (PSTU)	Ney Nunes (PCB)
19 de agosto de 2014	Band Rio	Sérgio Costa	Presente	Presente	Presente	Presente	Presente	Não convidado	Não convidado
2 de setembro de 2014	RedeTV!, Portal IG, Jornal O Dia	Amanda Klein	Presente	Presente	Presente	Presente	Presente	Não convidado	Não convidado
26 de setembro de 2014	Record Rio	Gustavo Marques	Presente	Presente	Presente	Presente	Presente	Não convidado	Não convidado
30 de setembro de 2014 ²³	TV Globo Rio de Janeiro	Ana Paula Araújo	Presente	Presente	Presente	Presente	Presente	Não convidado	Não convidado

Fonte: Wikipedia (2014)

Os relatórios gerados pela equipe de analistas de mídias sociais da campanha também visualizavam as comparações das ocorrências nas redes sociais de cada candidato durante os debates políticos, como mostra o quadro 19. A pergunta que norteava essa análise era: *Qual foi a quantidade de ocorrências sobre os candidatos?*

Quadro 19: Ocorrências X Candidatos

CANDIDATO	BAND	REDE TV	RECORD	GLOBO
X	5.364	3.052	3.843	19.740
Y	5.804	1.232	1.583	6.599
W	1.347	562	1.173	8.259
Z	3.583	846	703	7.314
H	2.505	759	550	3.678

Fonte: elaboração da autora a partir de informações da campanha

O Quadro 19 mostra o resultado das ocorrências dos candidatos em todos os debates realizados no primeiro turno da campanha eleitoral de 2014. De acordo com o relatório gerado pela equipe de mídia social, o candidato teve mais ocorrências nos debates da Rede TV (3.052), na Record (3,843) e Globo (19.740), ficando atrás de candidato B (5.804) somente no primeiro debate realizado pela TV Bandeirante. As ocorrências eram captadas tanto das redes como *Facebook* quanto *Twitter* e nessa análise quantitativa eram consideradas todos os tipos de ocorrências, tanto negativas quanto positivas a respeito dos candidatos.

Os relatórios buscavam responder também a seguinte pergunta: *Qual foi o alcance estimado das menções feitas sobre candidatos?*

No Quadro 20 é apresentado o alcance que os candidatos tiveram em cada debate político nas redes sociais digitais. Esse tipo de monitoramento tinha por objetivo dar uma dimensão do candidato e seus adversários, bem como informar a equipe de estratégia e inteligência sobre o desempenho do **candidato X**.

Quadro 20: Comparativo – Alcance X Candidatos

CANDIDATO	BAND	REDE TV	RECORD	GLOBO
X	8.804.182	15.801.888	21.216.661	26.153.159
Y	14.680.262	9.414.715	15.222.101	20.140.885
W	6.660.567	5.331.375	7.319.171	14.999.311
Z	3.819.477	3.606.363	8.456.893	17.875.335
H	3.595.018	2.253.923	3.476.716	5.291.986

Fonte: elaboração da autora a partir das informações da Campanha

6.7.5 Para que serviam os dados e informações na campanha

Os dados e informações vindas de diversos ambientes, principalmente das mídias sociais (*Facebook* e *Twitter*) do candidato serviam principalmente para saber o que os eleitores estavam falando sobre a campanha, sinalizavam os sentimentos em relação ao candidato e seus adversários.

De que forma estavam recebendo as ações do candidato e também para saber como estava a evolução dos demais candidatos concorrentes. Os dados coletados guiavam as ações de comunicação da campanha de forma integrada, desde as mensagens divulgadas nos meios de comunicação de massa, por exemplo, no Horário Gratuito da Propaganda Eleitoral (HGPE)

na TV e no Rádio, na elaboração do conteúdo do programa, até mensagens mais segmentadas divulgadas nas redes sociais da internet e aplicativos como o WhatsApp. Além disso, existia o fortalecimento da campanha que era feita pelos cabos eleitorais que faziam abordagens na rua ou de porta em porta, para observar mais de perto os sentimentos dos eleitores cariocas em relação ao candidato e seus adversários.

Segundo um dos profissionais responsáveis pela equipe de rua da campanha, os cabos eleitorais tinham como objetivo “ouvir os eleitores” e não falar sobre as propostas do candidato. A estratégia era adotada para buscar cada vez mais abastecer as bases de dados com informações vindas diretamente dos cidadãos/eleitores sobre diversos temas e também sobre as expectativas que se tinham com o próximo governo.

Na dinâmica de busca de informações e dados dos eleitores pelos cabos eleitorais foram usados *tablets*, como já citado, que tinham como finalidade integrar os dados dos coletores com os dados gerenciados pela equipe de Tecnologia de Informação (Ideia Inteligência). Os especialistas em TI controlavam a chegada das informações no sistema Alexandria diretamente da empresa. Segundo informações dos especialistas, somente nesta campanha foram usadas em torno de 900 *tablets* por cabos eleitorais, que faziam pesquisas de porta em porta em todo o Estado do Rio de Janeiro.

Os dados das pesquisas com grupos focais, pesquisas por telefone e pesquisa sobre intenções de votos também formava um banco de dados e ia para o Sistema Alexandria, onde eram armazenados, processados, cruzados e analisados. O sistema Alexandria corria listas de dados e a partir de cada região ou cidade do Rio de Janeiro era possível saber como estava a disputa. A partir disso, as ações de campanha se intensificam em determinadas regiões para garantir a presença segmentada da campanha.

Com o resultado dessas análises era possível também fazer anúncios segmentados para diferentes públicos nas mídias sociais, como foi observado na comunicação digital e nas ações nos sites de redes sociais dos candidatos.

6.7.6 As informações na campanha digital

O objetivo nesta parte da pesquisa é descrever como a equipe de comunicação da campanha (**candidato X**) utilizava-se das informações coletadas ao longo do processo eleitoral de 2014 nas ações de comunicação com os eleitores. Como não tinha como acompanhar todas as áreas da comunicação eleitoral, foi feita a escolha por centralizar as observações na campanha digital, que de certa forma acaba refletindo o que acontecia com a campanha nos demais meios de comunicação.

Apesar da campanha estar dividida por áreas e meios de comunicação, as equipes trabalhavam de forma integrada, buscando assim unificar a campanha. Isso reflete o que diz um dos profissionais da campanha.

“Trabalhamos todos de forma integrada para termos uma noção do todo e a partir disso pensar a segmentação e direcionar a comunicação na campanha” (CMD).

Isso era evidenciado nas atividades e ações das equipes de profissionais envolvidas na campanha. Por exemplo, a equipe de comunicação digital sabia de todas as ações que a equipe de rua estava fazendo, assim como sabia quais programas eleitorais iram ao ar no Rádio e na Televisão em cada dia de propaganda. Além disso, os profissionais acompanhavam todas as ações da agenda do candidato, que era gerenciada por uma equipe de jornalistas. Da mesma forma, todas as ações voltadas para os meios impressos eram de conhecimento da equipe digital.

Segundo um dos profissionais, essa forma de integralizar a campanha era importante, porque as redes sociais digitais sempre eram os reflexos dos acontecimentos do dia.

“A campanha sempre repercute nas redes sociais. Eleitores satisfeitos ou insatisfeitos com o candidato, seja por meio de uma reportagem, do Horário Eleitoral na Televisão ou no Rádio, ou mesmo numa carreta ou visita acabam por manifestar as suas satisfações ou insatisfações sobre o candidato. E nós temos que estar atentos a isso para poder continuar conduzindo a campanha de uma forma geral” (CCD).

O resultado dessas análises de dados e informações era também transferido para o banco de dados do Sistema Alexandria. Assim o resultado das ações de monitoramentos alimentava e ao mesmo tempo se utilizava desses resultados para as suas ações de comunicação digital. Porém, muitos profissionais faziam na campanha as análises mais qualitativas (a olho nú), principalmente sobre algum fato polêmico que envolvia a campanha ou o **candidato X**.

Todas as ações eram baseadas em informações e análise de dados que vinham da estrutura de *big data*, a partir de relatórios que eram gerados pelo sistema Alexandria com análises da equipe de inteligência da campanha e repassados as demais equipes¹²⁹.

Além das pesquisas oficiais, eram feitas pesquisas internas sobre intenções de votos e a partir dos resultados muitas ações de comunicação eram mantidas ou então reformuladas. A velocidade com que as informações chegavam à campanha obrigava os profissionais a tomarem decisões rápidas.

¹²⁹ Nenhuma decisão sobre a comunicação eleitoral era tomada sem o aval da coordenação de comunicação e marketing político da campanha, sem o conhecimento do Diretório Estadual do Partido, assessores políticos e o próprio candidato. Esse modelo de hierarquia é que prevalece na maioria das campanhas majoritárias, mesmo sendo consideradas campanhas de comunicação modernas.

Uma das principais funções da equipe de comunicação era acompanhar tudo que saia nos telejornais diários, nas redes sociais e em outros canais digitais sobre o candidato e a campanha. Os dados eram coletados diariamente e analisados pela equipe de inteligência de dados da campanha. Todos os dias aconteciam reuniões com os principais líderes de cada segmento da campanha para alinhar mensagens e ações do candidato a partir dos fatos e acontecimentos do dia, que envolvia diretamente o candidato.

A campanha se manteve presente em todas as redes sociais (Facebook, Twitter e Instagram, Youtube). Para cada rede a comunicação era usada de forma segmentada, levando em consideração segundo um dos profissionais da campanha (CMD) os seus usuários e a forma como se usavam esses canais. Mas também havia integração dos conteúdos usados nas mídias tradicionais como por exemplos nas propagandas eleitorais gratuita na Televisão e no Rádio.

As redes eram usadas de duas formas pela equipe de comunicação digital:

- (1) como estratégias de monitoramento de dados (já discutido anteriormente);**
- (2) como estratégias de divulgação das propostas do candidato.**

A respeito das redes sociais, o Facebook era considerado o carro chefe da campanha, mas mantinha perfis em todas as outras redes como Twitter e Instagram.

“O Facebook é praticamente uma identidade virtual de cada pessoa (não só dos candidatos) por ele congrega texto e foto além de ferramentas pagas que permitem rápida, ampla e segmentação na divulgação. Consideramos na campanha o Facebook como o Jornal Nacional. O Twitter serve como uma mídia para divulgar e acionar mais facilmente a imprensa, os especialistas e formadores de opinião, não os eleitores diretamente. Já o Instagram ainda é pouco utilizado pelos eleitores, é uma rede mais elitizada, mas ajuda a dar visibilidade na campanha” (CMD).

Os conteúdos postados eram diários conforme a média de posts em cada rede. No Facebook a atualização era feita em cima de oito posts; no Instagram eram postadas em média cinco atualizações; no Twitter eram em média 20 por dia e no WhatsApp em média três atualizações de conteúdos. Com essa média de conteúdos por dia percebe-se que a campanha tinha mais foco nas redes Twitter (20) seguido do Facebook (8).

No que se refere ao uso para divulgação da campanha, a equipe de comunicação digital tinha como objetivo tornar a figura do candidato conhecida. Por ter um perfil de político mais técnico, sua imagem era considerada pela equipe de marketing pouco divulgada e conhecida no âmbito estadual. Na imprensa era considerado um político sem grandes ostentações de mídia e propaganda. Na avaliação da equipe de comunicação, o **candidato X** era um desafio, porque não poderia mudar a sua essência nos programas e nos conteúdos de comunicação. Todas as ações foram sempre associadas o candidato a um político humilde e

simples, como o post publicado no seu perfil do Facebook: “*Sou do interior a minha essência é essa*”. A mensagem vinha acompanhada de uma foto do candidato. Já no Twitter as mensagens faziam referências às mudanças que estavam só começando. E de que o candidato tinha muitas propostas e muita coisa por fazer pelo Estado do Rio de Janeiro.

No *Twitter* o candidato mantinha 14,8 mil seguidores e suas atualizações sobre a campanha eram constantes. A campanha seguia as mesmas linhas de comunicação do que as adotadas no *Facebook*, porém levando-se em consideração que os textos deveriam ser mais enxutos e objetivos atentos aos 140 caracteres de texto e que o público que seguia o **candidato X** era constituído de formadores de opinião.

Assim como nos outros meios de comunicação, nas redes sociais as estratégias de comunicação eram centradas nas antecipações a futuras críticas dos adversários, através de indicação de que certas políticas se iniciaram e precisavam ter continuidade. O papel central dos meios usados para a comunicação do **candidato X** com os eleitores fluminenses era de que ainda era preciso fazer mais pelo Estado no sentido de diminuir a desigualdade social em áreas como educação, saúde, transporte, principalmente em segurança e moradia. A ênfase era dada nas obras que estavam sendo realizadas por todo o Estado e as suas contribuições para o Rio de Janeiro.

Os aplicativos para celulares também foram amplamente utilizados nas estratégias de comunicação da campanha do candidato. Para atingir os eleitores do estado do Rio de Janeiro, a campanha também se utilizou do aplicativo do WhatsApp¹³⁰ para enviar mensagens aos eleitores cariocas. O aplicativo WhatsApp é um aplicativo para celulares *smartphones* e *tablets*, que transmite mensagens com rapidez e eficiência. É um aplicativo que tem como principais usuários, os jovens.

Segundo um dos especialistas em mídias digitais, as comunicações eram também segmentadas e para cada grupo dos três mil seguidores, que utilizava o aplicativo eram enviadas mensagens diferenciadas, sempre de acordo com os interesses do grupo e da sua região (localidade) do Estado. Utilizando esses canais a equipe da campanha mandava pelo aplicativo vídeos nos quais o candidato fazia promessas de campanha baseados em temas e necessidades dos grupos de eleitores, previamente identificados e analisados pela equipe de

¹³⁰ Segundo informações do WhatsApp (2014) o aplicativo conta com 38 milhões de brasileiros que utilizam mensalmente o serviço de mensagem. O número de usuários equivale a mais de 8% da base de 465 milhões de usuários em todo mundo. Entre os países que se utilizam dos serviços estão a Índia com 40 milhões de usuários, o México com 32 milhões, a Espanha com 25 milhões, o Reino Unido com 17 milhões e 13 milhões de usuários na Argentina.

dados da campanha. Além disso, no vídeo o **candidato X** convidava o eleitor a visitar sua página no Facebook e conhecer seu plano para o futuro do estado.

No primeiro turno a preocupação da equipe de comunicação do candidato era fazer com que ele se tornasse conhecido e suas propostas fossem divulgadas, pequenos trechos dos vídeos que eram apresentados no Horário Eleitoral eram enviados também com esse propósito.

O analista de mídias digitais, responsável pela coordenação da comunicação digital da campanha conta que em cada segmento eram identificados os eleitores em potencial¹³¹ (já previamente identificados pela equipe de inteligência de dados a partir das técnicas de *microtargeting*), as classes sociais eram segmentadas em B, C, D e as mensagens de campanha eram direcionadas para esses grupos, porém de uma forma ainda muito macro. Segundo ele as pesquisas de identificação de eleitores do candidato X apontava que a zona sul do Rio de Janeiro não era um grupo favorável ao candidato, por isso, poucas estratégias de comunicação eram feitas nessa região. Os temas mais tratados na campanha para cada região eram assim distribuídos:

- Zona Sul do Rio: Segurança;
- Zona Norte: Mobilidade e Segurança;
- Zona Oeste: Mobilidade e Segurança;
- Baixada Fluminense: Segurança, Infraestrutura, Transporte;
- São Gonçalo: Segurança e Transporte;
- Interior: Desenvolvimento Econômico e Infraestrutura.

Dentro desses temas outros subtemas eram trabalhados como **BRTs, UPPs, Arco Metropolitano, UPAS, Bibliotecas Públicas**, entre outros.

6.8 CONSIDERAÇÕES SOBRE A PESQUISA DE CAMPO

A partir da pesquisa empírica realizada no comitê eleitoral, usando a metodologia da observação direta e os depoimentos colhidos dos profissionais envolvidos na campanha política, é possível tecer algumas considerações em relação ao fenômeno observado e aos dados coletados no período de realização da investigação. Mesmo a metodologia escolhida não seguir nenhum rigor teórico na sua aplicação, acredita-se na viabilidade da análise empreendida nessa pesquisa.

Em relação às possíveis **transformações nas estruturas de campanhas eleitorais com o uso do *big data*** percebe-se que os fatores determinantes das ações da campanha

¹³¹ Informações não liberadas pela equipe de comunicação.

sempre eram embasadas a partir das **pesquisas, dados, informações e interações estabelecidas com os eleitores.**

Mesmo não tendo acesso aos dados e análises elaboradas pela equipe de dados e pesquisa, foi possível observar pelo movimento das equipes que as ações de comunicação giravam em torno de informações mais precisas da campanha. Nenhum passo era dado, sem antes uma prévia análise e discussões com as equipes responsáveis pela coleta de informações e dados, bem como com as equipes responsáveis pelas estratégias de comunicação e marketing.

No que se refere à **segmentação da campanha**, foi possível observar que as estratégias de comunicação eram direcionadas para as classes B, C e D, distribuídas por regiões eleitorais no Estado todo. A classe A e a zona sul da cidade do Rio de Janeiro ficavam de fora das estratégias de comunicação direta, porque segundo os profissionais responsáveis pelas pesquisas da campanha, a partir do **mapeamento eleitoral** feito inicialmente, por meio da coleta de dados e análises do *big data*, os resultados apontavam uma baixa aceitação do **candidato X** nessa região e com esse tipo de eleitor. Porém, não foi possível observar em que nível de segmentação essas informações e mensagens chegavam até os eleitores fluminenses. Durante a fala de alguns profissionais que cuidavam da campanha digital o nível de detalhamento dos eleitores não era tão grande assim e que a campanha ainda se preocupava em segmentar os eleitores por classe, sexo, idade, localidade, escolaridade.

A pesquisa teve dificuldades em acessar os dados mais precisos dos relatórios oficiais da campanha, o que não permite afirmar e comprovar **que todos os eleitores classificados pelas análises de *big data* e segmentados pelas técnicas do *microtargeting* foram atingidos com a comunicação direta e com o marketing individualizado.** Durante a pesquisa só foi possível ao acesso às imagens retiradas do sistema Alexandria (apresentadas no corpo da pesquisa e representadas por figuras), com suas bases de dados vazias, que foi disponibilizada pela equipe de inteligência de dados. Isso serviu apenas para ilustrar e referenciar o trabalho que o sistema era capaz de fazer na campanha.

Por outro lado, evidenciou-se que a campanha buscava atingir os eleitores em **diversos meios de comunicação.** Isso se comprova pelas estratégias de comunicação digital empreendidas nas mídias sociais como Facebook e o Twitter, que foram as principais mídias digitais usadas não somente para divulgar a campanha, mas também **para coletar dados e informações** que conduziram parte das estratégias de marketing. A campanha também se distribuiu pelo WhatsApp, ainda usado de forma tímida, mas se aproveitou das interações com

os usuários para a captura de dados para a campanha. Nesse ambiente não foi possível realizar uma pesquisa mais aprofundada.

Durante a observação também foi possível identificar o **grau de profissionalização da campanha eleitoral e a integração das atividades voltadas para a captura de dados e informações e nos usos desses resultados para a tomada de decisões**. A campanha contava com profissionais que cuidavam desde a pesquisa de rua até o monitoramento de mídias sociais. Eram todos profissionais com experiências em campanhas eleitorais nacionais e internacionais, com especialidades específicas como Jornalismo, Publicidade e Propaganda, Analista de Mídias Digitais, Sociólogos, Cientistas Políticos.

A informação norteava todas as ações e procurava-se nos debates do candidato em programas de televisão que suas propostas e discussões fossem sempre embasadas a partir de dados e informações verdadeiras e de qualidade. Por isso, **bancos de dados foram criados para direcionar as ações do candidato**. A partir desse banco, a equipe de comunicação digital elaborava os posts para o Facebook e Twitter com informações e dados precisos sobre diferentes temas levantados, durante o debate eleitoral na televisão. Assim que eram escolhidos os temas pelo gerenciador do debate, a equipe de comunicação digital, integrada com a equipe de jornalistas, que cuidavam do banco de dados, elaboravam anúncios e posts com dados e propostas do candidato para aquele tema. Esses anúncios eram imediatamente postados nas mídias sociais, demonstrando a dinâmica da campanha e a integração dos profissionais envolvidos. A instantaneidade, bem como a interatividade eram características marcantes de toda a campanha.

Os relatórios diários com informações e dados vindos de diversas fontes como da rua, por telefone, por meio de pesquisas focais (que eram realizadas de 3 a 4 por semana), bem como dos monitoramentos das mídias sociais, conduziam todas as ações da equipe de comunicação. Dependendo dos resultados, principalmente de cada fato que envolvia o candidato na rua, nos debates na televisão, nos programas eleitorais, na imprensa ou mesmo na sua agenda pública, eram imediatamente medidos a partir dessas coletas de dados e informações. Assim, novas análises eram empreendidas sobre o estado da campanha, dando condições para que a coordenação de comunicação revisse suas estratégias de marketing. Isso vai de encontro com que Ferraz (2010) afirma que dentre todas as transformações ocorridas nas campanhas eleitorais a mais importante é a inclusão da pesquisa política, no seu núcleo de estratégias e decisão. **Assim, na campanha observada nenhuma ação de comunicação era feita antes de rever os índices de aceitação e rejeição dos**

candidatos, que tinham um tratamento mais específico a partir do uso do *big data*, que conseguia agrupar as informações por segmentos de eleitores.

No que se refere ao monitoramento e coleta de dados das mídias sociais, mesmo sendo uma prática usual no mercado para coletar informações sobre o comportamento de consumidores e eleitores, o que mais chamou atenção na pesquisa foi a capacidade que essas ações tinham de **gerenciar e direcionar as informações da campanha**. Isso foi percebido durante os debates políticos que aconteciam na televisão e que automaticamente se refletia nas redes sociais. A partir dos comentários e das menções de eleitores e usuários das redes sociais, envolvendo o processo eleitoral e o candidato, novas estratégias de comunicação eram empreendidas. **Ficando claro que a vigilância e o monitoramento era um ponto central em toda a campanha**. Isso comprova que com o fim do anonimato digital, a capacidade de vigiar cidadãos nas mídias sociais virou uma prática deliberada para ajudar a direcionar propagandas e anúncios com a ajuda de *softwares* especializados para controle do candidato/da campanha. No entanto, essa vigilância se caracterizava como o que Fernanda Bruno (2009) chama de **vigilância distribuída**, que se espalha pelos meios, com interesses diferentes, sem ter o sentido de punição (Foucault), mas de buscar conhecimentos sobre o comportamento eleitoral, traçando padrões de comportamento e com interesses de conduzir esses eleitores na decisão do voto a favor do candidato. Assim, **o monitoramento era entendido como forma de observação para acumular informações, visando projeções ou construção de novas estratégias para a campanha**. Dessa forma, comprova-se também que a tecnologia e a profissionalização em campanhas estão cada vez mais presentes, principalmente para aquelas que têm a preocupação de conduzir suas ações a partir dos sentimentos e opiniões dos eleitores, no sentido de ganhar vantagens em relação aos seus adversários.

Enfim, as observações empreendidas nessa campanha eleitoral demonstram a importância da coleta e análise de dados na condução de toda a campanha, porém evidencia que se não houver um propósito claro de como utilizar essas informações, selecionando mensagens para eleitores certos, de nada adiantará tanta pesquisa e dados numa campanha eleitoral.

7 CONSIDERAÇÕES FINAIS

Essa tese teve como propósito estimular o debate em relação aos usos do *big data* em campanhas eleitorais brasileiras. Três hipóteses nortearam o trabalho de pesquisa - (1) o uso do *big data* em campanhas eleitorais, inclusive no Brasil, muda as estruturas e formas de se fazerem campanhas; (2) o grande volume de informação gerado e compartilhado em ambientes digitais impõe novos olhares e estratégias para as campanhas eleitorais; (3) o *big data* tende a alterar as relações entre eleitores e candidatos, permitindo que a campanha tenha melhor conhecimento em relação ao comportamento dos eleitores, contribuindo para a segmentação eleitoral.- para responder aos problemas levantados nesse trabalho que se configuram em investigar em que medida o *big data* tem sido usado em campanhas eleitorais? E que tipos de mudanças podem ser percebidas nas campanhas eleitorais brasileiras com os usos do *big data*?

A partir das questões levantadas e considerando as hipóteses, é possível tecer algumas considerações sobre o objeto investigado. A literatura, apesar de não ser muito vasta principalmente em relação ao foco da pesquisa, ajudou a entender toda a dinâmica que sustenta os conceitos mais contemporâneos do estudo, ajudando no entendimento da prática do *big data* na campanha eleitoral brasileira, que foi observada na disputa eleitoral de 2014 para governo do Rio de Janeiro.

Como se viu durante a pesquisa de campo, a temática é considerada ainda pouco explorada no campo científico brasileiro, bem como pouco publicizada entre profissionais, que se dedicam a elaborar estratégias informacionais e comunicacionais em campanhas eleitorais, usando dessa metodologia. Nesse cenário, a dificuldade de comparações se fizerem presente e foram tidos como um desafio para o desenvolvimento dessa tese.

No entanto, chega-se a conclusões de que a informação se coloca como um elemento central das principais discussões quando se faz referências ao *big data*. Confirma-se que a preocupação com a informação não é algo apenas dos dias atuais ou de intenções específicas das produções e gerenciamento de campanhas eleitorais. Essa é uma preocupação da humanidade desde os tempos mais remotos. Coletar, guardar e analisar dados é uma prática vivenciada muito antes do aparecimento das tecnologias e do próprio estabelecimento do conceito de datificação da informação. Desde sempre administrações e governos buscaram mais conhecimentos sobre as pessoas e seus governados com múltiplas intencionalidades. O que muda hoje são as ferramentas e a tecnologia empregadas na busca por esses dados e informações, que torno o processo cada vez mais facilitado e necessário no uso estratégico.

Dessa forma, não é possível construir uma campanha eleitoral eficiente sem informações sobre o eleitorado a quem a campanha se destina. E essas informações tem que ser de qualidade e retratar a verdade sobre o cenário eleitoral do momento e as possíveis chance da candidatura. Informação de má qualidade, ou seja, imprecisas, incompletas, incorretas empurra a campanha para o perigoso território do “palpitisimo”, da intuição dos marqueteiros e do ativismo voluntarista.

Portanto, até aí nenhuma novidade em relação ao uso de dados eleitorais por meio de pesquisas na campanha eleitoral de 2014. Como foi apresentado durante o case da campanha do deputado federal Basílio Villani, desde a década de 1986 já usava a coleta e análise de dados em suas campanhas eleitorais, mesmo não contando com ferramentas e metodologias específicas dos dias atuais. Dessa forma, o candidato e sua equipe estão sempre fazendo estimativas a partir dos dados e informações da campanha, a fim de atingir segmentos específicos de eleitores.

Como afirmam os especialistas entrevistados, as campanhas de hoje requerem informações confiáveis e de qualidade para um diagnóstico mais preciso, inteligência competitiva para que se possa focar e posicionar com eficiência a candidatura.

Por outro lado, a pesquisa apontou por meio das entrevistas realizadas com especialistas, que a segmentação de eleitores no Brasil é ainda muito simplista, genérica e imprecisa, isso porque não existe um real investimento na coleta de informações metodologicamente sistematizada. Existem ainda as dificuldades de cruzamento de informações de várias fontes públicas e privadas, que venham tanto de ambientes on-line quanto off-line para se chegar a variáveis mais específicas sobre os eleitores brasileiros. Por isso, a necessidade de contratação de profissionais especializados em *big data* em campanhas que se preocupam com informações mais precisas sobre o eleitorado brasileiro.

Assim, a abordagem empreendida nesse estudo leva para o entendimento de que o uso do *big data* em campanhas eleitorais é considerado muito **mais estratégico** do que tecnológico. A tecnologia é apenas o suporte por onde os dados e informações serão tratados e analisados. No entanto, o **valor** dos dados não está no seu volume em si, mas nos *insights* que são gerados a partir de uma clara visão do que é **necessário analisar**. O termo *big data* nasce a partir da possibilidade de gerenciamento e usos estratégicos da informação encontrada em diferentes ambientes (online e off-line) e que se usados corretamente podem fazer a diferença numa campanha eleitoral considerada “apertada”, em que o voto a voto está sendo disputado entre o candidato e seus adversários.

Considera-se o *big data* como um ecossistema que se (re) alimenta constantemente de **grandes volumes** de informações e dados, que vem de **diferentes fontes** (pessoas e coisas), em **diferentes formatos** (estruturados e não estruturados), num **fluxo contínuo e acelerado**. Isso impõe aos pesquisadores da informação e da comunicação novos olhares para o entendimento dessas relações complexas.

Durante a pesquisa de campo constata-se que o grande volume de dados disponíveis hoje, principalmente em ambientes digitais, impõe novos olhares sobre a possibilidade de (re) usos estratégicos dos dados individuais e particulares dos eleitores, o que comprova a hipótese inicial. As empresas pesquisadas (Elife e Ideia Inteligência) aproveitam-se da possibilidade das visibilidades das informações individuais em vários ambientes, bem como das possibilidades técnicas de coleta de grandes volumes de dados, com o propósito de oferecer serviços focados nas necessidades de políticos e candidatos. Dessa forma, oferece profissionais qualificados e tecnologias apropriadas para a realização dos serviços de *big data* focado em campanhas eleitorais segmentadas. Esse tipo de serviço teve seu início nos processos eleitorais de 2010 e vem sendo gradativamente ampliados e estendidos em campanhas eleitorais brasileiras.

As entrevistas realizadas com especialistas também revelaram pontos importantes sobre as possibilidades de usos do *big data* em campanhas eleitorais. Um ponto comum entre a maioria dos entrevistados é de que o uso do *big data* se configura como algo positivo, mesmo que o uso de dados individuais, que sustenta a maior parte das bases de dados do *big data*, sejam considerados privados e particulares. Para eles, o uso dos grandes volumes de dados durante o processo traria mais vantagens aos candidatos, análises mais certeiras, contribuindo sobremaneira para a elaboração de estratégias de informação e comunicação junto ao eleitorado em potencial (segmentado), bem como poderia traduzir em custo-benefício na campanha, principalmente direcionando estratégias específicas de comunicação para cada tipo de segmento de eleitores.

Porém, como era de se esperar nem todas as campanhas brasileiras entendem a real importância do uso do *big data* e se mostram predispostas a investir nesse tipo de trabalho, que requer disponibilidade financeira para investimento em equipes especializadas e tecnologias específicas na coleta e análise dos grandes volumes de dados. Na visão dos especialistas, somente campanhas eleitorais com disponibilidade financeiras investem na centralidade de dados para conduzir suas ações de informação e comunicação eleitorais. Mesmo assim, as campanhas majoritárias (cargos executivos) ainda são conduzidas com a fortes interferências partidárias, se tornando muito mais políticas do que informacionais e

comunicacionais, o que de certa forma dificulta proferir mudanças significativas nas estratégias de informação e comunicação. Porém, as expectativas dos especialistas é de que na medida que o *big data* se popularize no campo político, os seus usos se ampliem também. Esse tipo de coleta de informação, processamento e análise, microsegmentação e comunicação por grupo de eleitores similares é o rumo das novas campanhas eleitorais.

No que diz respeito a campanha eleitoral observada, a disponibilidade de dados e informações em diferentes ambientes, principalmente em redes sociais digitais, ampliou durante a disputa eleitoral de 2014 as possibilidades de coleta, análise e gerenciamento dos dados eleitorais. As estratégias de informação e comunicação eram conduzidas a partir de relatórios de análise dos dados coletados em relação à campanha. Como observado a internet não somava-se apenas como novo canal de comunicação, seu grande benefício foi no poder de coleta de informação, segmentação dos destinatários da mensagem, e mensuração imediata do resultado de ação (posts, twitters, Facebook, etc.). O candidato pode enviar a mensagem específica para o eleitor segmentado em variáveis de acordo com interesses da campanha. Para chegar a esses eleitores por meios desses canais, a campanha se preocupou em detectar anteriormente, a partir de um mapa eleitoral, com análise de dados minuciosa feitos pelo sistema *big data* Alexandria.

Isso implicou na mudança de estrutura das campanhas eleitorais, incluindo profissionais especializados desde a coleta até a análise dos dados. A campanha apostou em gerentes de dados, analistas, economistas comportamentais, cientistas políticos e sociais e programadores, na elaboração de instrumentos e aplicativos para criar valor e oferecer novas bases para a tomada de decisão de estratégia e comunicação na campanha. No entanto, os especialistas em marketing eleitoral com larga experiência em campanhas eleitorais nacionais e internacionais não foram dispensados, tanto que todas as decisões finais passavam pela gerência de marketing da campanha, que era representado por um publicitário. Visualizou-se a preocupação com o uso de tecnologia e ferramentas específicas para facilitar as ações de coleta de dados, análises e elaboração de mensagens nos meios de comunicação.

Assim, a campanha eleitoral analisada se misturava entre técnicas e metodologias inovadoras a partir do *big data*, mas também se complementava com as antigas técnicas de marketing eleitoral, que se utilizava de ferramentas e meios de comunicação tradicionais para atingir o eleitorado desejado. As mudanças mais evidentes na campanha eleitoral estavam na estrutura física e profissional, que contava com qualificações específicas e experiências comprovadas, bem como na forma de **se tomar as decisões para as estratégias de**

comunicação, levando-se em consideração os dados levantados em diferentes tipos de pesquisas e no cruzamento possível desses dados.

No que se refere as estratégias de informação e comunicação usadas na campanha, o uso do *big data*, reacende as “**velhas**” **formas de se comunicar com os eleitores**, como é o caso da campanha porta a porta, dos contatos diretos da campanha com o eleitor, mas de forma mais direcionada, podendo a campanha ter a certeza de que a comunicação direta está alcançando o eleitor certo. Assim, as campanhas passam a ser customizadas e o eleitor se torna o alvo de todo o processo. Porém, mesmo que o marketing direto passe a ser o centro das estratégias de campanhas eleitorais, como observado na pesquisa empírica, a campanha de massa ainda tem seu núcleo central. A televisão para os especialistas em campanhas eleitorais, tem o seu lugar central nas campanhas eleitorais brasileiras. Porém, é na segmentação da comunicação, que a campanha vai encontrar a sua grande diferença em relação as demais campanhas.

No entanto, a disponibilidade de informações em ambientes digitais, a facilidade na coleta e análise desses dados oferecidos por ferramentas, tecnologias e profissionais especializados faz reacender o ponto crítico dos usos do *big data* em campanha eleitorais, que se configura no estímulo a vigilância, controle e monitoramento das ações dos eleitores. A pesquisa comprovou na campanha eleitoral observada que se usava de monitoramento das ações digitais dos eleitores nas principais redes sociais (Facebook e Twitter) de forma instrumental pelos analistas de mídias digitais.

Durante o período de observação dentro do comitê foi observado essa prática sendo feita 24 horas por dia, por profissionais especializados em rastrear os dados e informações das pessoas, bem como em gerenciar *software* de captura dessas informações, que eram feitas não somente sobre os eleitores, mas também sobre os candidatos adversários. Esses resultados, como apresentado no decorrer da pesquisa, direcionava parte das ações de marketing digital do candidato, bem como serviam para alimentar os bancos de dados do Sistema Alexandria, que reunia dados de diferentes bancos para fazer as análises mais gerais da campanha.

No entanto, seria uma atitude no mínimo ingênua da pesquisa ignorar que o ambiente das redes sociais digitais monitorados durante a campanha eleitoral observada estejam imunem a efeitos panópticos, ou seja, que os dados aí publicados não possam servir a setores específicos, fora a própria intenção instantânea da campanha, instaurando dinâmicas através das quais alguns coletam, manejam e colocam ao serviço de seus interesses as informações individuais disponíveis. Assim, nas redes sociais, a conectividade generalizada serve tanto aos objetivos sociais de uso do meio quanto a formas organizadas de vigilância de dados através

das quais poucos observam muitos, se valendo das habilidades computacionais aumentadas que subsidiam as políticas de coleta, registro e tratamento da informação.

Assim, chegada do *big data* na vida das pessoas deve ser discutido além dos perigos à privacidade, porque para isso temos leis que garantem de certa forma o direito à informação pessoal. O perigo, agora, não está apenas nos usos primários dos dados, mas sim nos seus usos secundários. Nas outras possibilidades que esses usos com outras finalidades podem levar. Algo percebido e questionado durante a pesquisa foi o que seriam feitos com os dados e informações dos eleitores fluminenses depois da campanha eleitoral de 2014? Essa questão não foi respondida na pesquisa, mas que pode ser focada em trabalhos futuros e que não podem ser desconsideradas no campo científico.

Por outro lado, entende-se com as pesquisas realizadas nessa tese que mesmo que se busquem exemplos e modelos a partir de outras campanhas, inclusive internacionais, como a de Barack Obama, cada campanha eleitoral vai encontrar a sua própria maneira de usar o *big data*, sem seguir uma rigidez metodológica ou um modelo único. Cada cenário eleitoral é um cenário diferenciado. O que pode ser aplicado a uma campanha, não significa que possa e deva ser aplicado a outra com a mesma exigência de eficácia. Por isso, não pode-se comprovar na campanha eleitoral observada nessa pesquisa, que o uso do *big data* efetivamente contribuiu para o sucesso da candidatura. Seriam necessárias análises mais detalhadas e precisas sobre essa possibilidade e que essa investigação não se propôs a fazer.

Dessa forma, a partir dessa pesquisa, conclui-se que *big data* no universo de campanhas eleitorais pode ser entendido como um método de pesquisa, que usa dos dados, da informação e da tecnologia aliado às estratégias de comunicação tradicionais para conquistar segmentos de eleitores em potencial.

Vale ressaltar que um dos grandes aprendizados trazidos por esse trabalho está no fato de entender que *big data* não se trata apenas dos dados digitais (como se imaginava no início da pesquisa), mas sim de **qualquer informação e dado**, que possa ser coletado, **principalmente aquele buscado diretamente na fonte**, no caso da pesquisa, no próprio eleitorado. E que qualquer tipo de ferramenta e metodologia de coleta de dados e informações, pode contribuir para esse conceito, assim como a possibilidades de relações podem ampliar essa coleta de informações mais precisas. Assim também, que o *big data* sozinho não faz “milagres” é preciso um investimento tecnológico e humano, pois demanda atenção à infraestrutura e tecnologias analíticas, bem como percepções dos cenários para onde e para quem as mensagens devam ser direcionadas.

No entanto, apesar da pesquisa apontar diferentes direções sobre os usos do *big data* em campanhas eleitorais, conclui-se esse trabalho com a certeza de que muitas lacunas ainda poderão ser preenchidas em relação a essa temática. Isso porque os usos do *big data* em campanhas eleitorais no Brasil estão na sua fase embrionária. Poucos profissionais de comunicação têm se dedicado a usar o *big data* em suas estratégias de marketing eleitoral e quando usam essas iniciativas, ficam mais no campo do sigilo, com pouco ou nenhum interesse em pronunciar suas experiências e compartilhamento com o campo científico.

Assim propõem-se diferentes visões de investigação, para que cada vez mais *big data* possa ser divulgado no âmbito da produção científica, em áreas como Ciência da Informação e da Comunicação. E, que também possa servir de referências para impulsionar o campo prático, despertando interesses de profissionais da informação e comunicação, que se dedicam às campanhas eleitorais.

REFERÊNCIAS

ALBAGLI, Sarita. **Informação, saber vivo e trabalho imaterial** In: ALBAGLI, Sarita (Org.). *Fronteiras da Ciência da Informação*. Brasília: IBICT, 2013. p.107-126. Disponível em: <http://livroaberto.ibict.br/bitstream/1/1020/8/Fronteiras%20da%20Ciencia%20da%20Informacao.pdf> Acesso em: 17 jun. 2015.

ALBAGLI, Sarita; MACIEL, Maria Lucia. **Informação, poder e política: a partir do Sul, para além do Sul**. In: MACIEL, Maria Lucia; ALBAGLI, Sarita. *Informação, Conhecimento e Poder: mudança tecnológica e inovação social*. Rio de Janeiro: Ed. Garamond, 2011. p. 9-39.

ALBAGLI, Sarita; MACIEL, Maria Lucia. *Novos rumos da interdisciplinaridade*. Liinc em Revista, Rio de Janeiro, v. 1, n. 1, p. 1-3, 2005.

ALBUQUERQUE, Afonso. **“Aqui você vê a verdade na tevê”**: a propaganda política na televisão. Niterói: MCII, 1999.

ALDÉ, Alessandra. *A construção da política: cidadão comum, mídia e atitude política*. 234 f. (Tese de Doutorado, Ciência Política). Instituto Universitário de Pesquisas do Rio de Janeiro – IUPERJ. Rio de Janeiro, 2001.

ALMEIDA, Jorge. **Marketing político, hegemonia e contra-hegemonia**. São Paulo, Editora Fundação Perseu Abramo e Xamã Editora, 2002.

ALMEIDA, Jorge. **Pesquisas de opinião e cenário de representação da política**. [S.P.:s.n.] 2001. Disponível em: <<http://www.unb.br/fac/comunicacaoepolitica/Jorge2001.pdf>> Acesso em: 26 maio 2008.

ALVAREZ, Maria Esmeralda Ballesteros. **Organização, Sistemas e Métodos**. São Paulo: McGraw Hill, 1991, v. 1 e 2.

AMARAL, S.A. do. *Marketing: abordagem em unidades de informação*. Brasília: Thesaurus, 1998.

ANÁLISE em larga escala de dados não convencionais e diferentes. 2013. Disponível em: <<http://www-03.ibm.com/software/products/br/pt/infobigbasiedit/>>. Acesso em: 06 mai. 2013.

ANHAIA, Valéria dos Santos. **O marketing de relacionamento na campanha eleitoral de Yeda Crusius**. Santa Maria – RS, 2007. Disponível em: <http://lapecpp.files.wordpress.com/2011/05/valc3a9ria-anhaia.pdf>. Acesso em: 20 de dezembro de 2014.

ANTONIUTTI, Cleide Luciane. **De operário a presidente. A imagem política de Lula no Horário Eleitoral da televisão nas campanhas presidências de 1989 e 2002**. Dissertação (Mestrado em Sociologia Política), UFPR, Curitiba – Pr, 2004. Disponível em:

http://www.educadores.diaadia.pr.gov.br/arquivos/File/2010/artigos_teses/SOCIOLOGIA/clei de.PDF. Acesso em: set. 2013.

AQUINO, Judith. “**Transforming Social Media Data and Predictive Analytics**” Issue17. Vol. 2. November, 2012. Disponível em: crm.com. Acesso: 25 Jan. 2013.

ARAUJO, Eliany Alvarenga de. A construção social da informação: dinâmicas e contextos. **Data Grama Zero – Revista de Ciência da Informação**, v. 2, n. 5, out. 2001. Disponível em: < http://dgz.org.br/dez08/F_I_aut.htm >. Acesso em: 2 set 2013.

ARAUJO, Vania Maria Rodrigues Hermes de. Informação: instrumento de dominação e de submissão. **Ciência da Informação**, v. 20, n. 1, 1991.

ARAUJO, Vania Maria Rodrigues Hermes de. Sistemas de recuperação da informação: nova abordagem teórico conceitual. Tese (Doutorado em Comunicação e Cultura), Escola de Comunicação da Universidade Federal do Rio de Janeiro, 1994.

ARAUJO, Vania Maria Rodrigues Hermes de; FREIRE, Isa Maria. A rede Internet como canal de comunicação, na perspectiva da Ciência da Informação. **Transinformação**, v.8, n.2, 1996.

ÁREA do Brasil. 2013. Disponível em: <<http://www.brasilecola.com/brasil/area-brasil.htm>>. Acesso em: 04 jun. 2013.

ARRIGONI, Rodrigo. **Big data, a ciência do grande**. 2013. Disponível em: <http://convergecom.com.br/tiinside/21/06/2d013/big-data-a-ciencia-do-grande/#.Uoy0-sRJ6s1>. Acesso em: 18 Abr. 2014.

AZEVEDO, Fernando. **Espaço público, mídia e modernização das campanhas eleitorais no Brasil**. In: VII REUNIÃO ANUAL DA COMPÓS. São Paulo, 1998: PUC, jul, 1998. Trabalho apresentado na VII Reunião anual da Compós.

BARABÁSI, Albert-László. **Linked: the new science of networks**. Kindle Edition. Cambridge: Perseu Books Group, 2002.

BARABÁSI, Albert-László. **Burst: the hidden patterns behind everything we do**. Kindle Edition. London: Dutton (Penguin Books Ltd), 2010. BARABÁSI, Albert-László. **Thinking in network terms**. Edge Conversation. Set. 2012. Disponível em <http://edge.org/conversation/thinking-in-network-terms>. Acesso em: 29 set. 2012.

BARBOSA, Suzana. **Banco de dados como metáfora para o jornalismo digital de terceira geração**. In: LUSOCOM, 6., 2004. Anais Lusocom. Covilhã, Portugal, 2004a. CDROM.

BARBOSA, Suzana. **Identificando remediações e rupturas no uso de bancos de dados no jornalismo digital**. In: SBPJor, 2., 2004. Anais do II SBPJor. SalvadorBA/Brasil, 2004b.

BARRETO, Aldo. A condição da informação. **São Paulo em Perspectiva**. 67-74, 2002. Disponível em: <http://www.scielo.br/pdf/spp/v16n3/13563.pdf>. Acesso em: 20 fev. 2015.

BARRETO, Aldo de Albuquerque . A oferta e a demanda da informação: condições técnicas, econômicas e políticas. **Ciência da Informação**, v.28, n.2, p.168-168, 1999b. Disponível em:http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-

19651999000200009&lng=en&nrm=iso. Acesso em: 18out.2013. *big data*. 2013. Disponível em: <<http://www.gartner.com/it-glossary/big-data/>>. Acesso em: 14 abr. 2013.

BARRETO, Aldo de Albuquerque. Mudança estrutural no fluxo do conhecimento: a comunicação eletrônica. **Ciência da Informação**, v. 27, n. 2, p. 122-127, maio/ago. 1998. Disponível em: <<http://dx.doi.org/10.1590/S0100-19651998000200003>>. Acesso em: 2 out 2014.

BARRETO, Aldo de Albuquerque. A questão da informação. **São Paulo em Perspectiva**, São Paulo, v. 8, n. 4, p. 3-8, 1994.

BASTOS, Celso Ribeiro; MARTINS, Ives Gandra. **Comentários à Constituição do Brasil**. São Paulo: Saraiva, v. 2, p. 63, 1989.

BAUMANN, Zygmunt. Modernidade líquida. Rio de Janeiro: Jorge Zahar, 2001.

BAUMANN, Zygmunt. Vigilância líquida. Rio de Janeiro: Jorge Zahar, 2013.

BEAL, Adriana. *Segurança da informação: princípios e as melhores práticas para a proteção dos ativos de informações nas organizações*. São Paulo: Atlas, 2008.

BEL, Daniel. O advento da sociedade industrial: uma tentativa de previsão social. São Paulo: Cultrix, 1978.

BLUMER, John G. KAVANAGH, Dennis. The third Age of political communication: influences and features. **Political Communication**, v.16, p.209-230. 198596.x, 1999. Doi: 10.1080/105846099.

BOOTH, Wayne C. **A arte da pesquisa**. São Paulo: Ed. Martins Fontes, 2000.

BORKO, Hilda. Information science: what is it? *American Documentation*, Jan. 1968.

BRANDÃO JÚNIOR, Francisco de Assis. **Palanques virtuais em construção**. http://repositorio.unb.br/bitstream/10482/1176/1/DISSERTACAO_2008_FranciscoDeAssisFBJunior.pdf2008. Disponível em: http://repositorio.unb.br/bitstream/10482/1176/1/DISSERTACAO_2008_FranciscoDeAssisFBJunior.pdf. Acesso em: 20 jul 2014.

BROOKES, B. C. The foundation of Information Science. **Journal of Information Science**, v. 2, Part I, p.125-133, Part II A equação fundamental da Ciência da Informação... Frederico César Mafrá Pereira Inf. Londrina, v.13, n1, p.15-31, jan./jun. 2008, p.209-221, Part III p.269-275, and v. 3, Part IV, p.3-12, 1980/1981.

BRUNO, Fernanda. **Blog visitado**. 2009. Disponível em: <http://dispositivodevisibilidade.blogspot.com.br/>. Acesso em: 06 jan 2014.

BRUNO, Fernanda. Contramanual para câmeas inteligentes: vigilância, tecnologia e percepção. *Galaxia*, São Paulo, On-line, n. 24, p. 47-63, dez 2012.

BRUNO, Fernanda. Dispositivos de vigilância no ciberespaço: duplos digitais e identidades simuladas. **Revista Fronteira**, São Leopoldo, RS, v. 8, p. 152-159, 2006.

BRUNO, Fernanda. Mapas do crime: vigilância distribuída e participação na cultura contemporânea. In: BRUNO, F; KANASHIRO, M; FIRMINO, R. **Vigilância e visibilidade: espaço, tecnologia e identificação**. Porto Alegre: Sulina, 2010.

BRUNO, Fernanda. Máquinas de ver, modos de ser: vigilância, tecnologia e subjetividade. Porto Alegre: Sulina, 2013.

BRYNJOLFSSON, Erik; MCAFEE, Andrew. **The second machine age work, progress, and prosperity in a time of brilliant Technologies**. 2014. Disponível em: <<https://tanguduavinash.files.wordpress.com/2014/02/the-second-machine-age-erik-brynjolfsson2.pdf>>. Acesso em: 30 mar 2015.

BUCKLAND, M.K. Information as thing. **Journal of the American Society for Information Science (JASIS)**, v.45, n.5, p.351-360, 1991.

BURKE, Peter. Uma história social do conhecimento: de Gutenberg a Diderot. Rio de Janeiro: Jorge Zahar, 2003. 241 p.

BURKE, Peter. Uma história social do conhecimento II. Da Enciclopédia à Wikipédia. Rio de Janeiro: Jorge Zahar, 2012. 414 p.

CALVO, Tereza Cristina Mate. **Distinção entre informação e conhecimento**. Disponível em:http://www.microkids.com.br/artigos/distincao_entre_informacao_e_conhecimento.pdf Acesso em: 5 Set de 2013.

CANAVILHAS, João. **A comunicação política na era da internet**. Corvilhã, Portugal, Universidade da Beira Interior, 2009. Disponível em: <http://www.bocc.ubi.pt/pag/canavilhas-joao-comunicacao-politica-na-era-da-Internet.pdf> . Acesso em: 06 set 2013.

CAPPELLA, Felipe. **As novas campanhas eleitorais**. Disponível em: <felipecapella.blogspot.com.br/2014/03/2014-as-novas-campanhas-eleitorais.html>, 2014. Acesso em: 05 Jan. 2015.

CAPURRO, Rafael. **Epistemologia e Ciência da Informação**. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 5, Belo Horizonte, 2003. **Anais...** Belo Horizonte: Escola de Ciência da informação da UFMG, 2003.

CAPURRO, Rafael; HJORLAND, Birger. O conceito de Informação. **Perspectivas em Ciência da Informação**, v.12, n.1, p.148-207, jan./abr. 2007. Disponível em: <http://www.scielo.br/pdf/pci/v12n1/11.pdf>. Acesso em: 28 fev. 2015.

CARVALHO, G. M. R.; TAVARES, M. S. Informação & conhecimento: uma abordagem organizacional. Rio de Janeiro: Qualitymark, 2001.

CASARTELLI, Alam de Oliveira. **Proposição de um modelo de comunicação e marketing político**. São Paulo: INTERCOM, v. 10, p. 10-11, 2005.

CASTELLS, Manuel. A sociedade em rede. São Paulo: Paz e Terra, 1999.

CHAIA, Vera. Um mago do marketing político. **Comunicação & Política**, v. 3 n. 3, 1996.

CHALEGRE, Ágata M.; ALMEIDA, Jocely S. C.; PASSOS, Rubislam O. A importância das pesquisas de opinião para a elaboração de uma estratégia eleitoral. **Diálogos & Ciência — Revista Eletrônica da Faculdade de Tecnologia e Ciências**. Ano 4, n. 7, jun. 2006. ISSN: 1678-0493. Disponível em: <http://www.ftc.br/revistafsa>. Acesso em: 25 Jan. 2014.

CHOO, Chun Wei. A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. São Paulo: Editora Senac, 2003.

CLARKE, Roger. Information technology and dataveillance. **Communications of the ACM**, v.37, p. 498–512, 1988.

CONSULTAS e artigos com a palavra “análise”. 2005. Disponível em: <http://origemdapalavra.com.br/palavras/analise/>. Acesso em: 19 abr. 2013.

CORREA, Manuel. Uso de Big Data ajuda governo brasileiro a gastar de forma mais eficiente. 2013. Disponível em: <http://oglobo.globo.com/tecnologia/uso-de-big-data-ajuda-governobrasileiro-gastar-de-forma-mais-eficiente-8582240>. Acesso em: 12 Out. 2013.

COSTA, Caio Tulio. O papel da Internet na conquista dos votos de Marina Silva, 2011. Disponível em: <http://interessenacional.uol.com.br/index.php/edicoes-revista/o-papel-da-internet-na-conquista-dos-votos-de-marina-silva/>. Acesso em: 10 Maio 2014.

COSTA, Luciana Sodr ; CAVALCANTI, Marcos. Big data n o   uma tecnologia. **DataGramZero - Revista de Informa o** - v.15 n.3 jun/14. Disponível em: http://www.dgz.org.br/jun14/Ind_com.htm#autor. Acesso em: 16 Out. 2014.

COSTA, Luciana Sodr ; FERNANDES, Soraya; ZUPPO, Daniella Big Data: estudo do ambiente, desafios e an lise estrat gica para o Brasil. Rio de Janeiro: UFRJ/COPPE, 2013.

COUTINHO. Marcelo O. Marketing eleitoral: para n o desperdi ar recursos. S o Paulo:  cone, 1998.

COX, Michael.; ELLSWORTH, David. **Application-controlled demand paging for out-of-core visualizatoin.** 2004. Disponível em: http://www.evl.uic.edu/cavern/rg/20040525_renambot/Viz/parallel_volviz/paging_outofcore_viz97.pdf. Acesso em: 12 mai. 2013.

CUKIER, Kenneth. Data, data, everywhere: a special report on managing information. **The Economist**, v. 394, n. 867, Feb. 2010.

DAHLBERG, Lincoln. Computer-mediated communication and the public sphere: a critical analysis. **Journal of Computer-Mediated Communication**, v. 7, n. 1, 2001.

DAL VECHIO, Cintia. Visibilidade, vigil ncia, identidade e indexa o: a quest o da privacidade nas redes sociais digitais. Logos 34. **O estatuto da cibercultura no Brasil**, v. 34, n.1, 1 sem. 2011.

DANTAS, Edmundo Brand o. **A informa o como insumo da pr tica do marketing na gest o do conhecimento do cliente.** Bras lia: Universidade de Bras lia, 2006. Disponível: <http://www.bocc.ubi.pt/pag/dantas-edmundo-informacao-pratica-marketing.pdf>. Acesso em: 25 Fev 2014.

DAVENPORT, Thomas; BARTH, Paul; BEAN, Randy. How 'big data' is different. **MIT Sloan Management Review**, 30, jul, 2012. Disponível em: <http://sloanreview.mit.edu/article/how-big-data-is-different/>. Acesso em: 02 Dez 2014.

DAVENPORT, Thomas; HARRIS, Jeanne. **Competição analítica: vencendo através da nova ciência**. Rio de Janeiro: Campus, 2007.

DAVIS, Gordon. B.; OLSON, M. H. **Sistemas de información gerencial**. Bogotá: McGraw-Hill, 1987.

DAYAN, Daniel; KATZ, Elihu. **Media events: the live broadcasting of history**. Cambridge: Harvard University Press, 1992.

DELEUZE, Gilles. **Post-scriptum sobre as sociedades de controle conversações: 1972-1990**. Rio de Janeiro: Ed. 34, 1992. p. 219-226.

DELEUZE, Gilles. **Pourparlers**. Paris: Les Éditions de Minuit, 1990.

DEMCHENKO, Yuri. **Defining extra Vs: Value and Veracity for big data 3V: Volume, Velocity, Variety**. 2013. Disponível em: <http://www.uazone.org/demch/worksinprogress/sne-2013-02-techreport-bdaf-draft02.pdf>, Acesso em: 03 fev. 2013.

DIEBOLD, Francis X.. **On the origin(s) and development of the term “Big Data”**. Philadelphia. Penn Institute For Economic Research, 2012.

DUMBRILL, Edd. **Big data now: current perspectives from O'Reilly Media**. 2012. Disponível em: <http://oreilly.com/data/radarreports/big-data-now-2012/thankyou.csp>. Acesso em: 11 mar. 2013.

EATON, Chris. **Understanding big data**. IBM. McGraw-Hill Osborne Media ©2011.Ebook. 2012.

FEIJÓ, Bruno Vieira. A revolução dos dados. **Revista Exame PME – Pequenas e Médias Empresas**, São Paulo, p. 30-43, set. 2013.

FERRAZ, Francisco. **Manual completo de campanha eleitoral**. São Paulo: L&M Editores, 2010. 406 p.

FIGUEIREDO, Carlos. **Técnicas de campanha eleitoral**. São Paulo: Ibrasa, 1986.

FIGUEIREDO, Ney Lima; FIGUEIREDO JUNIOR, José Rubens de Lima. **Como ganhar uma eleição: lições de campanha e marketing político**. São Paulo: Cultura Editores Associados, 1990.

FIGUEREIDO, R. (Org.). **Marketing político e persuasão eleitoral**. São Paulo: Fundação Konrad Adenauer, 2000.

FOUCAULT, Michel. **Microfísica do poder**. Tradução de Roberto Machado. 18ª Edição. Rio de Janeiro: Graal, 2003.

FOUCAULT, Michel. **Vigiar e punir**. Rio de Janeiro: Vozes, 1975 p. 262.

FOUCAULT, Michel. **Vigiar e punir: nascimento da prisão**. Tradução de Raquel Ramalhete, 41. ed. Petrópolis, RJ: Vozes, 2013.

FREIRE, Gustavo Henrique de Araujo. Ética da Informação: Conceitos, Abordagens, Aplicações. In: SIMPÓSIO BRASILEIRO DE ÉTICA DA INFORMAÇÃO. João Pessoa: Ideia (PB), 2010.

FUCHS, Christian. **An alternative view of privacy on Facebook, in Information**. 2011. Disponível em: www.mdpi.com/2078-2489/2/1/140/pdf, pp. 140-165. Acesso em: 16 Abr.2012.

GARTNER. **Big Data**. www.gartner.com, IT Glossary. Disponível em: <http://www.gartner.com/it-glossary/big-data/>. Acesso em: 27 mar. 2014.

GIARDELLI, Gil. **Marketing político 2.0**. São Paulo:Exame.com, 2012. Disponível em <http://exame.abril.com.br/rede-de-blogs/pessoas-do-seculo-21/2012/11/13/marketing-politico-2-0/>. Acesso em: 20 de Jan. 2013.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2008.

GOMES, Neuza Demartini. **Formas persuasivas de comunicação política**. Propaganda política e publicidade eleitoral. 3. ed. Porto Alegre. EDUPUCRS, 2004.

GOMES, Wilson. Participação política on-line: questões e hipóteses de trabalho. In: GOMES, W.; MAIA, R. C. M.; MARQUES, F. P. J. A. **Participação política e Internet**. Porto Alegre: Editora Sulina, 2011.

GOMES, Wilson. Propaganda política, ética e democracia. In: RUBIM, Antonio C (Org). **Mídia, eleições e democracia**. São Paulo: Editora Sulina, 1994. p. 53-90.

GOUVEIA, Sônia Mara. A informação como produto de alto valor no mundo dos negócios. **CRB-8 Digital**. São Paulo, v. 1, n. 1, p. 43-54, jul. 2008. Disponível em: www.crb8.org.br/ojs. Acesso em: 02 Mar. 2012.

GRANDI, Rodolfo; MARINS, Alexandre; FALCÃO, Eduardo. **Voto é marketing... o resto é política**. São Paulo: Loyola, 1992.

IASULAITIS, Sylvia. **Internet e campanhas eleitorais: experiências interativas nas cibercampanhas presidenciais do cone sul**. (Tese de Doutorado em Ciências Políticas). UFSCar, São Paulo, 2012.

IBM. Disponível: <http://www-01.ibm.com/software/data/bigdata/>. Acesso em: 13 Jan. 2013.

INTEL IT Center. **Big data 101: Unstructured Data Analytics**. 2012. Disponível em: <http://www.intel.com/content/www/us/en/big-data/unstructured-data-analytics-paper.html>. Acesso em: 4 Abr. 2013.

ISSENBERG, Sasha. **How President Obama's campaign used big data to rally individual voters**, Part 1. Disponível em: <http://www.technologyreview.com/featuredstory/508836/how-obama-used-big-data-to-rally-voters-part-1/> . Acesso em: 17 Nov. 2013.

ISSENBERG, Sasha. "**Nudge the Vote: How Behavioral Science is Remaking Politics.**" New York Times Sunday Magazine, October 31, 2010, page MM28.

ISSENBERG, Sasha. **The victory lab: the secret science of winning campaigns**. New York, NY: Crown Publishers, 212a.

ISSENBERG, Sasha. "**The death of the hunch.**" Slate, March 22, 2012b. Disponível em: www.slate.com/articles/news_and_politics/victory_lab/2012/05/obama_campaign_ads_how_the_analyst_institute_is_helping_him_hone_his_message.html. Acesso em: 20 Nov. 2014.

ISSENBERG, Sasha. "**How president Obama's campaign used big data to rally individual voters, Part 2.**" MIT Technology Review, December 17, 2012c. Disponível em: www.technologyreview.com/featuredstory/508851/how-obama-wrangled-data-to-winhis-second-term/. Acesso em: 20 Nov. 2014.

JOHNSON, David J. **Gestão de redes de conhecimento**. São Paulo: Editora Senac, 2011.

KOTLER, Philip. **Administração de marketing**. São Paulo : Atlas, 1975.

KOTLER, Philip. **Marketing para o Século XXI – como criar, conquistar e dominar mercados**. 3ª ed. São Paulo: Editora Futura, 1999.

KOTLER, Philip. **Marketing 3.0. As forças que estão definindo o novo marketing centrado no ser humano**. Rio de Janeiro: Elsevier, 2010.

KRISHNAMURTHY, Balachander. **On the leakage of personally identifiable information via on-line social networks**. WOSN'09, August 17, Barcelona, Spain. **2009**. Disponível em: < <http://conferences.sigcomm.org/sigcomm/2009/workshops/wosn/papers/p7.pdf> >. Acesso em: 12 Abr. 2014.

KUNTZ, Ronald. **Manual de campanha eleitoral: marketing político**. 9ªed. São Paulo: Global, 2002.

KUNTZ, Ronald A.; LUYTEN, Joseph M. **Marketing político: a eficiência a serviço do candidato**. São Paulo: Global, 1982.

LAS CASAS, Alexandre Luzzi. **Marketing de serviços**. 4ª edi. São Paulo: Atlas, 2006.

LANEY, Douglas. "**The importance of 'Big Data': a definition**". Gartner, 2012. Disponível em: <http://www.gartner.com/resid=2057415>. Acesso em: 20 Jan. 2014.

LASTORI, Airton. **MySQL e big data**. São Paulo, 2013. Disponível em: <<http://www.slideshare.net/MySQLBR/mysql-e-big-data>>. Acesso em: 8 mai. 2013.

LAZER, David. et al. **Computacional Ciências Sociais**. 2006. Disponível em: <http://www.sciencemag.org/search?author1=David+Lazer&sortspec=date&submit=Submit>. Acesso em: 12 Set. 2013.

LAZER, David et al. **Computation social science**. Science, Washington, v. 323, p. 721-724, 2009.

LE COADIC, Yves-François. **A Ciência da Informação**. Brasília: Briquet de Lemos/Livros, 1996.

LE MOS, André. **Cibercultura: tecnologia e vida social na cultura contemporânea**. 7. ed. Porto Alegre: Sulina, 2015. 295 p. (Coleção Cibercultura).

LE MOS, André. **Cibercultura e mobilidade: a era da conexão**. In: Congresso Brasileiro de Ciências da Comunicação. Rio de Janeiro, 2005. Disponível em: www.intercom.org.br/papers/nacionais/2005/resumos/r1465-1.pdf. Acesso em: 3 de Jun..2013.

LE MOS, André. **A comunicação das coisas: teoria ator-rede e cibercultura**. São Paulo: Annablume, 2013.

LETOUZÉ, Emmanuel. **Big data for development: challenges & opportunities**. UN Global Pulse, maio 2012.

LÉVY, Pierre. **Cibercultura**. São Paulo: 34, 1999.

LÉVY, Pierre. **A inteligência coletiva: por uma antropologia do ciberespaço**. 2. ed. Tradução de Luiz Paulo Rouanet. São Paulo: Loyola, 1999.

LÉVY, Pierre. **As tecnologias da inteligência: o futuro do pensamento na era da informática**. Rio de Janeiro : Editora 34, 1993. 208 p.

LÉVY, Pierre. **O que é virtual?**. Tradução de Paulo Neves. São Paulo: Editora 34, 1996.

LIMA, C. C. C.; MONTEIRO, R. L. Panorama brasileiro sobre a proteção de dados pessoais: discussão e análise comparada . **AtoZ: novas práticas em informação e conhecimento**, Curitiba, v. 2, n. 1, p. 60-76, jan./jun. 2013. Disponível em: <<http://www.atoz.ufpr.br>>. Acesso em: 10 Jan. 2015.

LIMA, Marcelo O. Coutinho de. **Marketing eleitoral: para não desperdiçar recursos** São Paulo : Ícone, 1988.

LIMA JUNIOR, Walter. Jornalismo computacional em função da “Era do Big Data”. *Líbero*. São Paulo v. 14, n. 28, p. 45-52, dez. 2011. Disponível em: <http://casperlibero.edu.br/wp-content/uploads/2014/05/1-Jornalismo-computacional.pdf>. Acesso em: 10 mai. 2013.

LITTLE JONH, Stephen W. Teoria da informação. **In: Fundamentos teóricos da comunicação humana**. Rio de Janeiro: Guanabara, [197-]. p. 152-161.

LIU, Bing. Sentiment Analysis and Subjectivity. **In: Handbook of Natural Language Processing**. Segunda Edição. 2010

LOJKINE, Jean. **Uma abordagem sistêmica da Revolução Informacional**. São Paulo: Edi. Corte , 2002.

LORENA, Karin Mestriner. **Qual o custo da comunicação fragmentada?**. 2013. Disponível em: <http://www.ibm.com/midmarket/br/pt/library_dicas_comunicacao.html>. Acesso em: 6 abr. 2013.

LYON, David. Introduction. In:_____ (Org). **Surveillance as social sorting: privacy, risk and digital discrimination**. London: Routledge, 2003. p. 1-9.

LYON, David. **Surveillance studies: an overview**. Cambridge: Polity Press, 2007. 243 p.

LYON, David. **The electronic eye: the rise of surveillance society**. Minneapolis: University of Minnesota Press, 1994. 270 p.

LYON, David (org.). **Theorizing surveillance: the panopticon and beyond**. Cullompton: Willan Publishing, 2006. 351 p.

MANCINI, P; SWANSON, David L. **Politics, media and modern democracy: an international study of innovations in electoral campaigning and their consequences**. London: Praeger, 1996.

MANHANELLI, Carlos Augusto. **Estratégias eleitorais: marketing político**. São Paulo: Summus, 1988.

MANHANELLI, Carlos Augusto. **Eleição é guerra**. São Paulo: Summus, 1992.

MANIN, Bernard. As metamorfoses do governo representativo. **Revista Brasileira de Ciências Sociais**, n. 29, p. 5-34, 1995.

MANIS, M. **Processos cognitivos**. São Paulo: Herder, 1973.

MANOVICH, Lev. **The language of new media**. Cambridge: MIT Press, 2002.

MANOVICH, Lev. **The Language of new media**. The Mit Press. Cambridge, Massachus. London, England, 2000.

MANOVICH, L.; HOCHMAN, N. **Zooming into an instagram city: reading the local through social media**. First Monday, Chicago, v.18, n7, 2013. Disponível em: <http://firstmonday.org/article/view/4711/3698>. Acesso em: 24 mar . 2014.

MANYIKA, James; CHUI, Michael; BROWN, Brad, et al. **Big data: the next frontier for innovation, competition, and productivity**. McKinsey Global Institute, 2011.

MANZINI, E. J. **Entrevista: definição e classificação**. Marília: Unesp, 2004. 4 transparência. P&b, 39 cm x 15 cm.

MATTINGLY, James. **Radar screens, astroturf and dirty work: a qualitative exploration of process in corporate political action**. **Business and Society Review**, v.111, n.2, p.193-221, 2006.

MARTELETTO, Regina. Análise de redes sociais aplicação nos estudos de transferência da informação. **Ciência da Informação**. Brasília, v. 30, n. 1, p. 71-81, jan./abr. 2001. Disponível em: <http://www.scielo.br/pdf/ci/v30n1/a09v30n1.pdf>. Acesso em: 3 Jan. 2014.

MARTELETTO, Regina . Redes e configurações de comunicação e informação: construindo um modelo interpretativo de análise para o estudo da questão do conhecimento na sociedade. *Investigación Bibliotecológica*, México, v. 14, n. 29, p. 69-94, jul./dic. 2000.

MASSUCHIN, Michele Goulart. **Quais as funções do Websites durante as campanhas eleitorais? As eleições brasileiras de 2012 e o cenário de Curitiba**. UFSCar. CONGRESO DE COMUNICACIÓN POLITICA ESTRATÉGIAS DE CAMPAÑA. 2, Toluca, 2013.

MAYER-SCHONBERGER, Viktor. Delete:the virtue of forgetting in the digital age. Princeton, Princeton University Press: 2009.

MCAFEE, A; BRYNJOLFSSON, E. Big data: the management revolution. **Harvard Business Review**, Oct. 2012. Disponível em: <https://hbr.org/2012/10/big-data-the-management-revolution/ar> . Acesso em: 12 abril de 2014.

MCKENNA, Regis. **Marketing de Relacionamento**. Rio de Janeiro: Editora Campus, 1993.

MCLUHAN, Marshall. **Os meios de comunicação como extensões do homem (Understanding Media)**. 12 ed. São Paulo: Cultrix, 2002.

MIGUEL, Luis Felipe. **Os meios de comunicação e a prática política**. Lua Nova n. 55-56, p.155-184. 2002. Disponível em: <http://www.scielo.br/pdf/ln/n55-56/a07n5556.pdf>. Acesso em: 10 set. 2013.

MIKSA, F. L. Library and information science: two paradigms. In: VAKKARI, P.; CRONIN, B. (Eds.) **Conceptions of library and information science; historical empirical and theoretical perspectives**. London: Taylor and Graham, 1992. p. 229-243.

MILLER, Jacques-Alain. A máquina panóptica de Jeremy Bentham. Tradução de M.B. Magno. In: TADEU, Tomaz (Org.). **O Panóptico**. Traduções de Guacira Lopes Louro. M. D. Magno e Tomaz. Tadeu. 2. Ed. Belo Horizonte: Autêntico, 2008, p. 89-124.

MIRANDA, R. C. da R. O uso da informação na formulação de ações estratégicas pelas empresas. *Ciência da Informação*, Brasília, v.28, n.3, p.284-290, set./dez. 1999.

MORIN, Edgar. **O problema epistemológico da complexidade**. Lisboa: Mem Martins: Europa-América, 1996.

NAFFAH, S. **Análise Confiável**. [s.l.:s.n.] 2012. Disponível em: <http://www.sas.com/offices/latinamerica/brazil/sascom/revistas/pdfs/sascom_2012q2-ed-19.pdf>. Acesso em: 19 abr. 2013.

NEGREIROS. Davys Sleman. Eleições midiaticizadas: a modernização das campanhas. Davys Sleman de Negreiros Libros Básicos en la Historia del Campo Iberoamericano de Estudios en Comunicación. N. 75, fev -abr, 2011.

NEGROPONTE, Nicholas. **A vida digital**. 2. ed. Rio de Janeiro: Briquet de Lemos, 2006.

NICKERSON, David; TODD, Rogers. **Political campaigns and big data**. HKS Working Paper, 2014. No. RWP13-045. Available at SSRN: <http://ssrn.com/abstract=2354474> or <http://dx.doi.org/10.2139/ssrn.2354474>. Acesso em: 10 Abr. 2015.

NONAKA, Ikujiro; TAKEUCHI, Hirotaka. **Criação do conhecimento na empresa: como as empresas geram a dinâmica da inovação**. Rio de Janeiro: Campus, 1997.

NONAKA, Ikujiro, KROGH, Georg Von e ICHIJO, Kazuo. **Facilitando a Criação de Conhecimento**. Rio de Janeiro: Campus, 2001.

NORRIS, Pippa. **Democratic divide? The impact of the Internet on parliaments worldwide**. Havard: Harvard University, John Kennedy School of Government, 2000.

NORRIS, Pippa. **Digital divide; civic engagement, information poverty, and the Internet worldwide**. Cambridge: Cambridge University Press, 2001.

O ESTADO DE SÃO PAULO. **Big data**. Disponível em: <http://www.estadao.com.br/>. Acesso em: 14 mai. 2014.

O GLOBO. **Como funciona o big data**. Disponível em: <http://oglobo.globo.com/infograficos/bigdata/> Acesso em: 12 mai 2013.

O GLOBO. **As redes sociais e as eleições**. Disponível em: <http://noblat.oglobo.globo.com/artigos/noticia/2014/09/redes-sociais-e-eleicoes-por-gaudencio-torquato.html>. Acesso em: 4 mai 2014.

OHL, Rodolfo. **Big data: como analisar informações com qualidade**. Disponível em: <http://corporate.canaltech.com.br/coluna/big-data/Big-Data-como-analisar-informacoes-com-qualidade/>. Acesso em: 2 mai. 2015.

OLIVEIRA, M.; BERTUCCI, M. G. E. S. A pequena e média empresa e a gestão da informação. **Informação e Sociedade**, João Pessoa, v.13, n.2, p. 1-16, 2003. Disponível em: <http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/91>. Acesso em: 12 jan. 2013.

OLOFSON, Carl W; VILLARS, Richard L.; EASTWOOD, Mathew. **Big data: What it is and why you should care**. Paper, IDC Analyze the Future, 2011. Disponível em: www.idc.com. Acesso em: 30 Out 2013.

ORWELL, George. **1984**. Tradução de Alexandre Hubner, Heloisa Jahn. São Paulo: Companhia das Letras, 2009.

PACHECO, Cid. Marketing eleitoral: a política apolítica. In: **Comunicação e Política**. Rio de Janeiro: Ed. Cebela. Ano I, n 1, p. 147 / 153, 1994.

PANKE, Luciana et al. **O que os candidatos à Presidência do Brasil falaram nos programas do HGPE, nas últimas eleições? [S.l.:s.n.], 2011**. Disponível em: <http://www.compolitica.org/home/wp-content/uploads/2013/05/GT-09-Propaganda-e-Marketing-Pol%C3%ADtico-Luciana-Panke.pdf>. Acesso em 14.fev.2013.

PARISIER, Eli. **O filtro invisível: o que a internet está escondendo de você**. Rio de Janeiro: Zahar, 2012.

PEPPERS, Don; ROGERS, Martha. **Marketing um a um**: marketing individualizado na era do cliente. Rio de Janeiro: Campus, 1994.

PIMENTA, Ricardo M. Big data e controle da informação na era digital: tecnogênese de uma memória a serviço do mercado e do estado. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 6, n. 2, jul./dez. 2013.

POPKIN, Samuel. **The reasoning voter**: communication and persuasion in presidential campaigns. Chicago, London: Chicago University Press, 1996.

POSPIECH, Marco; FELDEN, Carten. Big Data – A state – of-the-art. In: AMERICA CONFERENCE ON INFORMATION SYSTEMS (AMCIS), 18, Seattle. **Proceedings...** Seattle, 2012.

POSTMAN, Neil. **Informing ourselves to death. German Informatics Society** (Gessellschaft fuer Informatik), Stuttgart, 11 Out. 1990. Disponível em: <http://w2.eff.org/Net_culture/Criticisms/informing_ourselves_to_death.paper>. Acesso em: 15 Nov 2014.

Entrevistas

BROWN, Donald. Diretor do Instituto Big Data da Universidade de Virgínia dos Estados Unidos. Entrevista realizada no dia 16 de dezembro de 2013.

CAVALCANTI, Marcos. Doutor em informática e professor da Coppe/UFRJ. Entrevista realizada no dia 05 de abril de 2015.

COSTA, Luciana S. Presidente do Instituto Big Data Brasil. Entrevista realizada no dia 27 de fevereiro de 2015.

GUERREIROS, Christian C. Professor e especialista da área da ciência da computação. Entrevista realizada no dia 26 de fevereiro de 2015.

ISSENBERG, Sasha. Jornalista especialista em Campanhas eleitorais americanas. Entrevista realizada no dia 19 de maio de 2015

LIMA JÚNIOR, Walter T. Professor e pesquisador da Universidade Metodista de São Paulo. Entrevista realizada no dia 25 de julho de 2014

MANHANELLI, Carlos. Especialista em Marketing Político e Campanhas Eleitorais, Professor e autor de livros. Entrevista realizada no dia 25 de 2014.

MOURA, Maurício J. S. B. Especialista em Campanhas Eleitorais no Brasil e nos Estados Unidos. Entrevista realizada no dia 23 de fevereiro de 2015.

PAULA, Carolina A. Dra em Ciência Política, coordenadora de pesquisa da empresa Ideia Inteligencia. Entrevista realizada no dia 24 de julho de 2014.

PEREIRA, Renato. Antropólogo e Publicitário, coordenador de campanhas políticas no Brasil e no exterior. Entrevista realizada no dia 10 de junho de 2014.

GUERREIROS, Christian C. Professor e especialista da área da ciência da computação. Entrevista realizada no dia 26 de fevereiro de 2015.

SEIXAS, Tatiana. Gerente de Marketing do Ideia Inteligência. Entrevista realizada no dia 04 de julho de 2014.

SIQUEIRA, Geraldo. Ex-assessor parlamentar do deputado Federal Basílio Villani (PSDB). Entrevista realizada no dia 01 de maio de 2015.

SOTTO, Breno. Gerente de Projetos da E.Life. Entrevista realizada por e-mail no dia 01 de junho de 2015.

TAURION, Cezar. Especialista em tecnologia da informação. Entrevista realizada no dia 04 de agosto de 2014.

VILLANI, Basílio. Ex-deputado federal do Paraná (PSDB). Entrevista realizada no dia 20 de maio de 2015.

APÊNDICE

APÊNDICE A – ROTEIRO DE ENTREVISTAS COM EMPRESAS**ROTEIRO DE ENTREVISTA
COM EMPRESA DE PESQUISA E DADOS**

- 1- Desde quando a empresa existe?
- 2- Quantos funcionários trabalham aqui?
- 3- Quais as especialidades?
- 4- Quem são os clientes da empresa?
- 5- Que tipos de atividades exercem?
- 6- Como usam o *big data*?
- 7- Qual é a tecnologia usada para o *big data*?
- 8- Como coletam os dados para as atividades que desenvolvem?
- 9- De onde vêm os dados que compõem o sistema *big data*?
- 10- Quantas campanhas eleitorais já trabalham oferecendo serviços de *big data*?

APÊNDICE B – ROTEIRO DE ENTREVISTA COM PROFISSIONAIS**ROTEIRO DE ENTREVISTA****PROFISSIONAIS ESPECIALISTAS EM CAMPANHAS ELEITORAIS**

1. Como você avalia o uso do *big data* em campanhas eleitorais?
2. Já teve experiências com o *big data* fora do cenário brasileiro? Quais? Como foram?
2. Da sua experiência na campanha de Obama de 2012, o que pode ser aproveitado na realidade brasileira nas campanhas de 2014? Quais foram as principais diferenças?
3. Das campanhas brasileiras que usaram o *big data* o quanto você credita o sucesso do candidato baseado no seu uso? Poderia justificar?
4. No que se aplica a técnica de *microtargeting* em relação ao *big data* numa campanha eleitoral?
5. Que plataformas foram usadas para o *big data* nas campanhas brasileiras?
6. Que tipos de inovações o *big data* trouxe ou poderá ainda trazer para as campanhas eleitorais e para a comunicação e marketing político?
7. Poderia destacar um dos momentos mais importantes da campanha eleitoral (primeiro e segundo turnos) em que as análises de dados foram usados, comprovando-se a eficiência do *big data*?
- 8- Que tipos de conhecimentos sobre os eleitores brasileiros foram possíveis com o uso do *big data* nas campanhas eleitorais de 2014?
- 9- Tudo que foi possível medir sobre os eleitores brasileiros foram medidos com o *big data* na campanha de 2014? Poderia dar alguns exemplos.
- 10 – do ponto de vista estratégico o que de importante o uso do *big data* poderá trazer para uma campanha política, principalmente brasileira?
- 11- Quais os riscos que o *big data* pode trazer?
- 12- Do ponto de vista do eleitor que implicações são resultantes do *big data*?

APENDICE C- ROTEIRO DE ENTREVISTAS COM ESPECIALISTAS**ROTEIRO DE ENTREVISTA****ESPECIALISTAS EM MARKETING POLÍTICO E ELEITORAL BRASILEIRO**

1. Como o marketing político brasileiro vê o uso do *big data* em campanhas políticas?
2. Que experiências já tivemos aqui no Brasil com o *big data*? e como foram em relação as campanhas que não utilizaram? demonstraram certas vantagens?
3. Podemos dizer que a campanha de obama de 2012 de certa forma inaugura uma tendência na forma de se fazer marketing político eleitoral a partir do uso do *big data*? por quê?
4. Quais são os pontos comuns nas experiências no uso do *big data* no âmbito brasileiro e em outros países? no que estas experiências se diferenciam?
5. O uso de *big data* provoca mudanças nas campanhas eleitorais? De que tipo?
6. O uso de *big data* em campanhas eleitorais tende a aumentar? Por quê? em que direção?
7. Como tem sido o uso de *big data* nas campanhas eleitorais no Brasil? quais as perspectivas futuras?
8. Candidatos que se utilizam do *big data* levam mais vantagens competitivas? por quê? quais?
9. No que se refere ao uso da internet, como as redes sociais têm sido usadas por candidatos/políticos em campanhas eleitorais? Existem um uso inteligente para as análises de dados (*big data*)? De que forma?

APENDICE D- ROTEIRO DE ENTREVISTA COM CIENTISTA POLÍTICA**ROTEIRO DE ENTREVISTA
CIENTISTA POLÍTICA**

- 1) Como define *big data* no contexto da política eleitoral?
- 2) O que de importante para as campanhas eleitorais brasileiras este conceito tem trazido?
- 3) Existe uma evolução do uso do *big data* aqui no Brasil? A partir de quando podemos observar isso? E em que direção isso está?
- 4) Existe algum case brasileiro importante no uso do *big data* em campanhas eleitorais? Qual? Como se deu?
- 5) Teve alguma experiência com o *big data* no Brasil? Quando? Como foi?
- 6) O que difere as campanhas eleitorais brasileiras com o uso do *big data*, da campanha eleitoral de Barack Obama de 2012? O que podemos aprender com esse caso?
- 7) Que mudanças substanciais o uso do *big data* traz para as campanhas políticas, principalmente brasileiras?
- 8) A partir do *big data*, existem mudanças na forma de se pensar a política e de se fazer campanhas eleitorais? Quais?
- 9) Que *insights* para as campanhas políticas podem ser extraídos?
- 10) Como seria a estrutura organizacional e profissional de uma campanha eleitoral baseada no uso do *big data*?
- 11) Quais profissionais na área de campanha política são necessários para trabalhar com o *big data*?
- 12) Em relação ao uso de dados pessoais de eleitores para compor as análises de *big data* como fica a questão ética e de privacidade?
- 13) Candidatos que se utilizam do *big data* levam mais vantagens competitivas? Por quê? Quais? Do ponto de vista da campanha (candidato) e do ponto de vista do eleitor.
- 14) Do ponto de vista da comunicação política que mudanças podemos ver em relação ao uso do *big data*? A comunicação mudou com os eleitores, de que forma?
- 15) Com o uso do *big data* as pesquisas eleitorais tradicionais passam por alguma transformação? Quais?
- 16) O que considera de negativo no uso do *big data* em campanhas eleitorais?

APÊNDICE E - DIÁRIO DE CAMPO

Este texto tem por objetivo relatar as vivências e experiências pessoais, profissionais e acadêmicas durante o trabalho realizado para doutoramento no Programa de Pós-Graduação em Ciência da Informação pelo IBICT/ECO-UFRJ. Os momentos vivenciados foram árduos, dadas as muitas dificuldades que foram sendo encontradas na caminhada da pesquisa. Além de escolher uma temática pouca explorada no campo científico, com poucas bibliografias para a construção do referencial teórico e poucos profissionais e especialistas na área da temática dispostos a contribuir, tive que lidar com situações pessoais, que muitas vezes me colocaram em desequilíbrio e desânimo para o desenvolvimento da tese.

No meio do caminho, quando já estava finalizando os créditos e disciplinas exigidas do programa (PPGCI) e dava início as pesquisas e leituras bibliográficas, após o nascimento do meu segundo filho (Agosto de 2012) minha mãe adoeceu desenvolvendo uma doença autoimune chamada de Guillan Barré¹³². Ficou por mais de dois meses hospitalizada, parte desse tempo, ficou na UTI em um hospital na cidade de Ponta Grossa, no Estado do Paraná. Em decorrência disso, ainda cuidando de um filho recém-nascido e com mais uma filha pequena de 4 anos, deixei por um tempo a pesquisa de lado e fui cuidar da minha família e ajudar no restabelecimento da saúde da minha mãe. Foram meses para começar a ver os primeiros resultados dos tratamentos e ainda entre o período da recuperação da minha mãe tivemos mais duas perdas familiares, a primeira foi a morte precoce da minha sobrinha de 22 anos em um acidente de carro fatal, no dia em que defendia o seu Trabalho de Conclusão de Curso em Enfermagem na UEPG e meses depois foi a morte do meu irmão de criação em consequência de uma cirrose hepática. Não precisa nem relatar como eu me encontrava emocionalmente diante de tudo isso.

¹³² A síndrome de Guillain-Barré, também conhecida por polirradiculoneuropatia idiopática aguda ou polirradiculopatia aguda imunomediada, é uma doença do sistema nervoso (neuropatia) adquirida, provavelmente de caráter autoimune, marcada pela perda da bainha de mielina e dos reflexos tendinosos. Ela se manifesta sob a forma de inflamação aguda desses nervos e, às vezes, das raízes nervosas, e pode afetar pessoas de qualquer idade, especialmente, os adultos mais velhos. O processo inflamatório e desmielizante interfere na condução do estímulo nervoso até os músculos e, em parte dos casos, no sentido contrário, isto é, na condução dos estímulos sensoriais até o cérebro. Em geral, a doença evolui rapidamente, atinge o ponto máximo de gravidade por volta da segunda ou terceira semana e regride devagar. Por isso, pode levar meses até o paciente ser considerado completamente curado. Em alguns casos, a doença pode tornar-se crônica ou recidivar. Site Drauzio Varella. Disponível em: <http://drauziovarella.com.br/letras/g/sindrome-de-guillain-barre/>. Acesso em: 12 Dez 2015.

Mesmo com muita tristeza, busquei nos meus filhos a força que precisava para a retomada da pesquisa de doutoramento. Assim, a minha pesquisa só foi retomada em março de 2013.

Vale lembrar que essa pesquisa foi reformulada. O projeto de pesquisa no qual fui aprovada no processo de seleção foi abortado no final de 2012, quando comecei a perceber que muitas pesquisas e trabalhos científicos vinham sendo desenvolvidos com o foco em redes sociais e comunicação política. A nova temática então (re)nasceu a partir de uma conversa inicial com a amiga e professora do curso de Comunicação Social das Universidade Federal do Paraná (UFPR), Dra Claudia Irene Quadros, com quem tive o prazer de trabalhar por muitos anos, ministrando aulas no curso de Comunicação Social (Jornalismo, Publicidade e Propaganda e Rádio e TV) na Universidade Tuiuti do Paraná (UTP). Cláudia Quadros foi minha coordenadora de curso e uma das grandes incentivadoras no entendimento da comunicação digital, naquela época (2000) em que a disciplina começava a ser ministrada por mim na referida instituição. Depois de muito conversar com ela e trocar ideias a respeito de temáticas interessantes para ser objeto de estudo de uma tese, chegou-se ao tema *big data*. E, como não queria me distanciar muito do campo da pesquisa em que já vinha desenvolvendo desde 2002 no mestrado em Sociologia Política na Universidade Federal do Paraná (UFPR), na linha de pesquisa em comunicação política, com foco em campanhas eleitorais, procurei um recorte do objeto no mesmo campo.

Além da influência sofrida a partir da conversa com a professora Claudia Quadros, a escolha pela temática *big data* em campanhas eleitorais, foi incentivada pelo case da campanha do presidente americano Barack Obama em 2012, que usou da tecnologia e das ferramentas e metodologias para a análise de grandes volumes de dados (*big data*), principalmente digitais, na construção das suas estratégias de comunicação. E também, pelo fato de ter trabalhado muitos anos como assessora de comunicação e marketing na área política, com o deputado federal Basílio Villani (na época PSDB) e ter gerenciado a sua campanha eleitoral de 1998.

A partir desse incentivo, as pesquisas começaram a ser centradas na busca por discussões teóricas e práticas nesse campo. Um dos principais canais de pesquisas usados durante essa investigação foi a própria internet, a partir dos sites de pesquisas como Google e plataformas de pesquisas científicas como periódicos Capes, Scielo, repositórios em bibliotecas de Universidades nacionais e internacionais, entre outros. Nos primeiros quatro meses de 2013 passei a me dedicar em pesquisar bibliografias e documentos sobre o case de Obama e a temática *big data*, a fim de entender o seu conceito e funcionamento. Além disso,

passei a manter contatos com grupos de estudos e de discussões em redes sociais (Facebook) que se dedicavam ao assunto. Na ânsia de encontrar documentos, artigos e reportagens sobre a campanha de Obama, passei a manter contato com profissionais americanos que tinham alguma proximidade e envolvimento profissional ou científico. Muitos emails foram encaminhados tentando essa aproximação, porém poucos se mostraram interessados em ajudar na divulgação de informações ou compartilhamento de seus estudos.

Diante da falta de respostas por parte desses profissionais e especialistas, passei a me sentir insegura em relação a temática escolhida para investigação e os questionamentos a respeito da viabilidade da pesquisa começaram a ser colocadas em discussão. Então, novas estratégias de pesquisa foram necessárias ao trabalho, direcionando o trabalho para a busca de profissionais e pesquisadores brasileiros que pudessem falar sobre suas experiências em campanhas eleitorais com o *big data*. O primeiro entrevistado foi o publicitário Renato Pereira que também é dono da Agência Prole, que está localizada no bairro da Urca (Rio de Janeiro). A primeira conversa com a intenção de entrevista rendeu bons frutos à pesquisa. Descobri que sua experiência em campanhas eleitorais com uso de *big data* era representativa tanto em processos eleitorais brasileiros quanto internacionais. E que ao contrário do que eu imaginava, campanhas brasileiras já usavam as análises de grandes volumes de dados desde 2010 em suas estratégias de comunicação e marketing eleitoral.

A partir desse primeiro contato foi possível chegar a outros profissionais que já usavam ferramentas e metodologias nas análises de grandes volumes de dados em campanhas eleitorais, como o cientista de dados Maurício Moura. Além de contribuir para a pesquisa com suas visões de especialista, Moura também abriu as portas da sua empresa no Rio de Janeiro, a Ideia Inteligência para que fosse realizada uma outra etapa da investigação do objeto, em que investigava como empresas especializadas em trabalhar com grandes volumes de dados (*big data*) vinham se posicionando frente ao mercado e ao segmento político eleitoral.

Depois de selecionada as duas empresas de pesquisa e dados - Elife de São Paulo e Ideia Inteligência do Rio de Janeiro - e mantive contato respectivamente com seus diretores, por meio de e-mails, dando início as primeiras entrevistas com os profissionais ligados as atividades das empresas. Tanto a empresa de São Paulo quanto a empresa do Rio de Janeiro não fizeram nenhum tipo de objeção a minha presença ou trabalho de pesquisa sobre as atividades que exerciam na área de *big data* e política eleitoral. As primeiras conversas se deram com a equipe da empresa Elife de São Paulo no dia 28 de abril de 2014, quando procurei saber quem poderiam ser meus entrevistados na empresa. Primeiro contato foi por

email e depois realizei uma pequena entrevista com um dos profissionais responsável pela gestão de projetos da empresa via telefone, no dia 27 de maio de 2014. Depois disso, ainda mantive mais contatos utilizando entrevistas por email com o proprietário da empresa e demais profissionais. As entrevistas com essa empresa se derão principalmente por telefone e email. Nenhuma visita pessoal e física foi feita a empresa em São Paulo por questões de agendas dos representantes. Toda informação da empresa para a pesquisa era encaminhada por meio de e-mails.

Já na empresa Ideia Inteligência, além das conversas mantidas por telefone e email, também fiz visitas ao local. No total foram quatro visitas feitas ao escritório da empresa que fica na rua São José, 90, sala 503, no centro da cidade do Rio. A empresa tem extensões também em Brasília, São Paulo, Portugal e nos Estados Unidos. Entre as visitas estão: uma no dia 16 de julho de 2014 em que conheci o lugar e parte dos profissionais que trabalhavam lá, fiz no momento uma entrevista com a cientista política Carolina de Paula. Outra no dia 24 de julho de 2014 em que comecei a fazer entrevistas com profissionais que cuidavam da parte tecnológica e para finalizar a primeira etapa das entrevistas antes da campanha eleitoral, realizei uma última visita no dia 28 de agosto de 2014 em que conversei com profissionais de comunicação e marketing. Para fechamento da pesquisa com a empresa, a última entrevista foi realizada no dia 30 de janeiro de 2015, com o objetivo de sanar as dúvidas levantadas durante a campanha eleitoral. Em todas as visitas percebi que se tratava de uma empresa pequena, que ocupava apenas dois andares de um prédio central. Num total de 10 funcionários apenas, distribuídos entre funções de tecnologia, pesquisa e coleta de dados eleitorais, estatística e marketing.

Paralelamente as pesquisas de campo que desenvolvia junto as duas empresas de dados, também fazia as entrevistas com especialistas e pesquisadores da área de *big data* no Brasil e no exterior. As entrevistas se deram em vários momentos como consta na parte de pesquisa de campo no corpo do trabalho. As entrevistas foram feitas de três maneiras: pessoalmente (face a face), via e-mail e Skype. Tanto pessoalmente quanto por Skype, as entrevistas tiveram em média a duração de 30 minutos a uma hora e foram gravadas, posteriormente transcritas e arquivadas, para que fossem consultadas sempre que houvesse necessidade no decorrer da escrita da tese. A entrevista partiu de um roteiro semiestruturado entre 8 a 12 questões abertas, dependendo da especialidade do profissional. As entrevistas foram realizadas de dezembro de 2013 a abril de 2015. Todos os entrevistados foram identificados mediante autorização. Foram entrevistados um total de 12 especialistas, sendo

10 brasileiros e 2 americanos. Essas entrevistas totalizaram 350 minutos de depoimentos gravados (Skype e pessoalmente).

Ao fazer contatos com Renato Pereira e Maurício Moura fiquei sabendo que eles iriam gerenciar a campanha de 2014 para governo do Estado do Rio de Janeiro e que usariam entre as suas estratégias de campanha as ferramentas e metodologias do *big data*. Nesse momento vislumbrei a possibilidade de me inserir como observadora dessa campanha. Mediante a autorização dos gerentes de comunicação e informação da campanha passei a fazer o trabalho de campo em agosto de 2014 diretamente no comitê político do candidato (prefiro deixar no anonimato). Para poder realizar a pesquisa no comitê tive que assinar um termo de confidencialidade sobre a não divulgação do nome do candidato, que para Renato Pereira e Maurício Moura é considerado como um cliente.

O comitê ficava localizado no bairro da Barra/Recreio do Bandeirantes, que fica num bairro nobre do Rio de Janeiro. Para chegar até o local eu gastava em média com transporte público umas duas horas, pegando dois ônibus e um metrô. Do local em que moro dá aproximadamente uns 30 quilômetros de distância. Tudo isso tinha que ser calculado por mim pela questão do tempo e trajeto. Nas primeiras duas semanas fiz esse trajeto todo de transporte público, mas a partir da terceira semana, já conhecendo melhor a rota e com auxílio de um GPS passei a usar o meu próprio carro, já que o local em que ficava o comitê oferecia um amplo espaço com estacionamento para os profissionais que estavam trabalhando na campanha.

O espaço em que ficava o comitê eleitoral era também chamado de polo, porque ali fora das disputas eleitorais era um espaço designado para locações de filmagens e produções audiovisuais. O prédio tinha a estrutura de um grande barracão, não sei precisar exatamente o tamanho em metros quadrados, mas era um local amplo e com o teto alto para facilitar qualquer instalação de equipamento e cabeamento de fios, sem muitas divisórias, sem janelas e com apenas uma porta de entrada e uma saída de emergência. Para entrar no local tinha que passar por uma mini recepção com uma catraca, somente quem tinha cartão de identificação tinham passagem livre. Eu por exemplo tinha que sempre estar acompanhada de alguém da equipe ou pedir autorização para as recepcionistas e seguranças do local, dizendo o porque da minha presença ali.

Nesse comitê ficavam as equipes de profissionais que cuidavam da comunicação e do marketing do candidato. Como era um espaço amplo, não existiam divisórias entre os profissionais. Ali internamente trabalhavam em torno de uns 300 profissionais para atender as ações de informação, comunicação e marketing do candidato. Além de profissionais

auxiliares, como recepcionistas, seguranças, faxineiras, pessoal que cuidava inclusive da logística e da contabilidade da campanha, totalizando mais de 1000 colaboradores. As áreas do comitê eram divididas por nichos, formando em cada posição do barracão um conjunto de mesas, com computadores e luminárias, que a noite eram usadas para iluminar as mesas de forma individual. Além desses equipamentos, tinha uma televisão central, em que os profissionais de jornalismo e publicidade acompanhavam os telejornais e os programas eleitorais, bem como um grande telão que transmitia para toda a equipe os programas eleitorais e os debates dos candidatos. No fundo do barracão, porém num lugar central ficava a mesa do gerente da campanha, responsável pela comunicação e marketing do candidato (Renato Pereira) e bem ao centro existia uma espécie de sala de estar, onde sempre que o candidato ou assessores do partido ficavam para discutir o andamento da campanha. Além disso, tinha uma sala de reuniões improvisada separadas por divisórias, local em que apenas se reuniam os representantes e gerentes de cada área da comunicação para discutir estratégias e avaliar peças publicitárias. As reuniões eram feitas frequentemente sempre que alguma coisa movimentava a campanha do candidato. Em anexo a esse espaço ainda existia os estúdios onde eram gravados os programas eleitorais para a televisão e para o rádio.

Para a realização da pesquisa, ficou acordado com a gerência de marketing da campanha, que a minha entrada no comitê seria condicionada em dias e horários. Então passei a visitar o local três vezes por semana – segundas, quartas e sextas – nos horários das 11h00 – 20h00. Em dias esporádicos eu fazia um período mais prolongado para acompanhar por exemplos os debates dos candidatos nas emissoras de televisão e observar como os profissionais de informação e comunicação se portavam na construção das suas estratégias, a partir do uso dos dados coletados nas redes sociais. Nesses dias o trabalho de observação ia até a finalização dos trabalhos dos profissionais, por volta de uma hora da manhã.

Eu não tinha um lugar exato para ficar no comitê, mas pela proximidade desde o início na empresa que era responsável pelas pesquisas e dados da campanha, passei a sentar bem próximos a eles para entender como se dava a dinâmica do trabalho da equipe. Até porque esses profissionais eram que me passavam a maior parte das informações sobre como era o andamento dos trabalhos ali no comitê. Depois dos primeiros dias de contato com alguns profissionais, vi que era necessário direcionar o olhar em alguma estratégia de comunicação. Voltei-me então para a comunicação digital, por considerar pertinente já que o embasamento teórico da minha tese fazia referências a isso.

Nas observações com equipe de comunicação digital foi possível ver como eram feitos os monitoramentos das redes sociais, a dinâmica e a instantaneidade da comunicação digital,

bem como a integração entre as mídias e como agiam os profissionais de comunicação na elaboração de estratégias do candidato. A equipe de comunicação digital, foi a equipe com quem mais me identifiquei na pesquisa. Sempre fui bem recebida. Todos que faziam parte da equipe tinham disposição para me ajudar e colaborar com os meus questionamentos, como para me mostrar o funcionamento de *softwares* que faziam a captura e análise dos dados nas redes sociais. Depois de uma semana de observação no comitê já me sentia mais próxima e familiarizada tanto da equipe de comunicação digital quanto da equipe de produção jornalística, com quem sempre conversava informalmente na intenção de tirar alguma informação sobre a campanha, quando não conseguia observar isso claramente ou por não estar presente no momento que isso acontecia no comitê.

Durante o período da minha pesquisa, consegui observar a movimentação da equipe com o candidato que fazia visitas constantes ao comitê, juntamente com assessores e lideranças do partido, para participar de reuniões com a gerência da campanha e aprovação de materiais e peças publicitárias da campanha.

Apesar de ter experiências em campanhas eleitorais, principalmente para cargos majoritários (deputado), com essa pesquisa foi possível ver o quanto é diferente uma campanha para governo de estado numa cidade como o Rio de Janeiro. A preocupação vai desde a estrutura, a disponibilidade de tecnologia até a especialidades dos profissionais envolvidos. Campanhas de grandes proporções precisam de muito investimentos tanto financeiros quanto profissionais, para que suas ações sejam eficientes.

AS LIMITAÇÕES DA PESQUISA

Durante a investigação me deparei com várias limitações na pesquisa que devem ser ressaltadas. A primeira foi a falta de literatura sobre a temática. A maioria encontrada e que serviram como orientação para a pesquisa colocam *big data* sempre como um tema ainda em construção, emergente, flexível e plausível de confusões conceituais, bem como com visões positivas em relação aos seus benefícios. As publicações começam a surgir a partir de 2012, momento em que essa tese também começa a ser delineada, mas ao longo dos últimos dois anos, inúmeras publicações sobre *big data*, surgiram ligados a negócios e tecnologia, que se caracterizaram como uma visão parcial do ponto de vista da efetividade amplitude dos usos do *big data*.

No âmbito acadêmico não encontrei praticamente nenhuma publicação na área de ciências sociais que contemplasse a temática com foco no campo das campanhas eleitorais, daí a dificuldade de estabelecer um diálogo maior com outras produções científicas. Essa falta

de publicações no caso brasileiro pode ser explicada porque poucas campanhas eleitorais na prática têm trabalhado com essa temática e os profissionais que usam dessas estratégias acabam por não revelar e divulgar de forma clara, informações que consideram restritas à campanha e ao candidato. Acabam por manter no sigilo de suas ações estratégicas, o que de certa forma dificulta o entendimento mais claro de como isso pode ser usado a favor da campanha e do candidato, e que resultados efetivos podem realmente trazer para aqueles candidatos que se utilizam dos recursos do *big data* para as suas estratégias de campanha.

Outras dificuldades encontradas em relação a escolha pela observação como técnica e métodos para a coleta de dados merecem ser destacadas. O cenário de uma campanha eleitoral vista pelo lado da produção da campanha é sempre muito intensa, principalmente pelo pouco tempo em que as campanhas precisam construir a imagem do candidato e colocar a campanha na rua e nos meios de comunicação. Não existe horário fixo para o desenvolvimento de determinadas ações, elas partes das necessidades imediatas da campanha de formular, rever e avaliar o impacto dela junto ao eleitorado. Essa dinâmica tende a ser cada vez mais intensa e fervorosa na medida em que as tecnologias da informação e comunicação proporcionam instantaneidade nas ações e reações da campanha eleitoral. E acompanhar isso não foi uma tarefa fácil. Muitas vezes, me pegava dispersa ao foco da observação que deveria ser empreendida, já que o planejamento da campanha se dava num único espaço, envolvendo diversas atividades de informação e comunicação e contava com muitos profissionais, circulando o tempo todo no mesmo espaço.

No que se refere às entrevistas realizadas com os especialistas tive que lidar com o fator tempo e falta de interesse de divulgar as experiências com a temática por parte de alguns entrevistados. Inclusive um dos entrevistados na pesquisa levou aproximadamente seis meses para me atender. E isso só aconteceu depois de inúmeros e contatos diretos feitos por telefone. Em um e-mail ele chega a ironizar dizendo: “*Se for por insistência em me entrevistar, você já é doutora*”. Somente para esse profissional eu encaminhei em torno de 10 *e-mails* para agendamento da entrevista, quando finalmente conseguia agendar, ele sempre desmarcava logo em seguida. Depois de inúmeras tentativas a entrevista que estava marcada para acontecer em outubro de 2014 só foi realizada no início de 2015.

Outra limitação encontrada foi em relação ao acesso a documentos e dados mais reais sobre a campanha eleitoral observada. Apesar de ter conseguido aprovação do gerente da campanha para realizar a minha pesquisa de observação direta, não pude contar com dados brutos e nem os relatórios finais da equipe de inteligência de dados sobre as análises dos eleitores fluminenses. Esses relatórios circulavam entre os profissionais e a *staff* da campanha,

mas em nenhum momento consegui ter acesso a eles. A pesquisa foi baseada praticamente nas observações, anotações e nos depoimentos dados pelos profissionais, a partir de uma visão mais otimistas e satisfatórias sobre os trabalhos, mas que não invalidam a pesquisa.

Além disso, as entrevistas com os profissionais envolvidos diretamente no processo desde a coleta até o uso dos dados foram orientadas pela coordenação da inteligência de dados da campanha, a fim de que informações consideradas confidenciais e sigilosas não fossem liberadas para a minha pesquisa. Isso de certa forma comprometeu o resultado da pesquisa no sentido de demonstrar a partir de dados reais como se dá o uso do *big data* em campanhas eleitorais. Por motivos óbvios e alheios muitos profissionais do comitê que procurei para entrevistar não quiseram responder as entrevistas, ficavam apreensivos em dar depoimentos ou mesmo em liberar informações da campanha, que julgavam ser confidenciais.

Outras limitações foram impostas à pesquisa como o período de permissão para o trabalho de observação. Minha permanência foi limitada aos meses de setembro e início de outubro, ou seja, apenas foi possível realizar a pesquisa em relação ao primeiro turno e depois que a equipe toda já estava montada e instalada no comitê. No segundo turno, por recomendações do partido, a gerência da campanha pediu para que eu não frequentasse mais o comitê político. Além disso, ficou estabelecido que a observação não seria feita de forma diária e em tempo integral, enquanto toda a equipe estivesse reunida, principalmente nas decisões mais estratégicas. Também, não consegui liberações de informações mais precisas sobre que tipos de grupos de eleitores a campanha buscava conquistar, alegando que isso eram informações confidenciais e que somente com autorizações do partido político do candidato, isso poderia ser liberado. Porém, todas as tentativas de contato com o presidente estadual do partido não foram atendidas, mesmo depois de finalizado o processo eleitoral.

Essas foram as minhas experiências de campo. Cabe ressaltar que em uma pesquisa científica, por mais que os caminhos a serem atingidos sejam traçados e organizados, o retrato que temos da realidade é sempre parcial, ainda mais quando se parte de uma pesquisa qualitativa e exploratória, com o uso de métodos e técnicas de observação. Por isso, a necessidade de se realizar outras pesquisas que tenha como temática o *big data*, mas com outros focos de investigações.