UNIVERSIDADE DE SÃO PAULO
ESCOLA DE COMUNICAÇÕES E ARTES
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA COMUNICAÇÃO

MAYANNA ESTEVANIM

Processos no jornalismo digital
do Big Data à visualização de dados

São Paulo
2016

MAYANNA ESTEVANIM

V.1

Processos no jornalismo digital

do Big Data à visualização de dados

Dissertação apresentada à Escola de Comunicações e Artes da Universidade de São Paulo (ECA/USP) para obtenção do título de Mestre em Ciências da Comunicação.

Área de Concentração: Teoria e Pesquisa em Comunicação. Linha de Pesquisa: Comunicação e Ambiências em Redes Digitais.

Orientadora: Profa. Dra. Elizabeth Nicolau Saad Corrêa.

São Paulo

2016

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na Publicação!

Serviço de Biblioteca e Documentação!

Escola de Comunicações e Artes da Universidade de São Paulo

Dados fornecidos pelo(a) autor(a)

Estevanim, Mayanna

Processos no jornalismo digital: do Big Data à
visualização de dados / Mayanna Estevanim. -- São Paulo: M.
Estevanim, 2016.

208 p.: il.

Dissertação (Mestrado) - Programa de Pós-Graduação em
Ciências da Comunicação - Escola de Comunicações e Artes /
Universidade de São Paulo.

Orientadora: Elizabeth Nicolau Saad Corrêa

Bibliografia

1. visualização de dados 2. sistema narrativo 3. Big Data
4. jornalismo de dados 5. jornalismo brasileiro I. Nicolau
Saad Corrêa, Elizabeth II. Título.

CDD 21.ed. - 302.2

Folha de aprovação

ESTEVANIM, Mayanna

PROCESSOS NO JORNALISMO DIGITAL: do Big Data à Visualização de Dados

Dissertação apresentada à Escola de Comunicações e Artes da Universidade de São Paulo (ECA/USP) para obtenção do título de Mestra em Ciências da Comunicação.

Área de Concentração: Teoria e Pesquisa em Comunicação

Linha de Pesquisa: Comunicação e Ambiências em Redes Digitais

Orientadora: Profa. Dra. Elizabeth Nicolau Saad Corrêa

Aprovado em: ____/____/____

Banca Examinadora

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Dedicatória

Para Eno, Gael, Maitê, Bella e há todos que defendem um jornalismo de qualidade.

Agradecimentos

À minha família –

Ao Eno, pelo companheirismo na caminhada; ao Gael pela paciência e a Maitê por ter auxiliado neste processo final. Aos meus pais Cristina e Carlos, irmão Leandro e cunhadas do coração. Aos demais familiares e amigos por compreenderem as ausências nos últimos tempos.

À minha orientadora –

À professora Elizabeth Nicolau Saad Côrrea pelo incentivo, dedicação, paciência e confiança depositados em mim nesta jornada acadêmica.

Aos professores e pesquisadores –

Aos pesquisadores do grupo COM+ da ECA-USP pelas discussões e trocas que auxiliaram nesta pesquisa. À professora Daniela Ramos por permitir que durante um ano fosse sua aluna PAE. Aos professores que generosamente aceitaram avaliar esse trabalho: na qualificação aos professores Daniela Bertocchi e Eugênio Bucci.

Aos apoiadores –

CNPq pela bolsa de estudos, à Paulo Carreta e Ana Patrícia Santos pelo companheirismo de estudos.

Agradeço a todos que participaram da construção e formação desta dissertação e pesquisadora.

“(...) tudo parece estar mudando ao mesmo tempo, e não existe um ponto privilegiado, acima da confusão, de onde eu possa enxergar as coisas”.
(Henry Jenkins, 2009, p. 39)

Resumo

ESTEVANIM, Mayanna. **Processos no jornalismo digital: do Big Data à visualização de dados**. Dissertação (mestrado). Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2016.

A sociedade está cada vez mais digitalizada, dados em diferentes extensões são passíveis de serem armazenados e correlacionados, temos um volume, variedade e velocidade de dados humanamente imensuráveis sem o auxílio de computadores. Neste cenário, falamos de um **jornalismo de dados** que visa o entendimento de temas complexos de relevância social e que sintoniza a profissão com as novas necessidades de compreensão informativa contemporânea. O intuito desta dissertação é problematizar a visualização de dados no **jornalismo brasileiro** partindo exatamente do que é esta **visualização de dados jornalísticos** e diante dos apontamentos sobre seu conceito e prática questionar como proporciona diferenciais relevantes. Por relevantes entendemos pautas de interesse público, que envolvem maior criticidade, maior aprofundamento e contextualização dos conteúdos no **Big Data**. As iniciativas que reúnem imagens relacionadas a dados e metadados ocorrem nas práticas de mercado, laboratórios acadêmicos, assim como em mídias independentes. Neste **sistema narrativo** atuam diferentes atores humanos e não-humanos, em construções iniciadas em codificações maquínicas, com bases de dados que dialogam com outras camadas até chegar a uma interface com o usuário. Há a necessidade de novas expertises por parte dos profissionais, trabalhos em equipe e conhecimento básico, muitas vezes, em linguagem de programação, estatística e a operacionalização de ferramentas na construção de narrativas dinâmicas e que cada vez mais envolvam o leitor. Sendo importante o pensar sobre um conteúdo que seja disponível para diferentes formatos. Para o desenvolvimento da pesquisa foi adotada uma estratégia multimetodológica, tendo os pressupostos da centralidade da comunicação, que perpassa todas as atividades comunicativas e informativas de forma transversal, sejam elas analógicas ou não. Um olhar que requer resiliências diante das abordagens teórico-metodológicas para que as mesmas consigam abarcar e sustentar as reflexões referentes ao dinâmico campo de estudos. Para se fazer as proposições e interpretações adotou-se como base o paradigma Jornalismo Digital em Base de Dados, tendo as contribuições dos conceitos de formato (RAMOS, 2012 e MACHADO, 2003), de jornalismo pós-industrial (COSTA, 2014), sistema narrativo e antenarrativa (BERTOCCHI, 2013) como meios de amadurecimento da compreensão do objeto proposto.

Palavras-chave: visualização de dados, sistema narrativo, jornalismo brasileiro, Big Data, jornalismo de dados

Abstract

ESTEVANIM, Mayanna. **Processes in digital journalism: from Big Data to data visualization.** Dissertation (Master's degree). Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2016.

Society is increasingly digitalized. Different scopes of data are likely to be stored and correlated, having volumes, variety and accumulating speeds humanly impossible to track and analyze without the aid of computers. In this scenario we explore the realm of **data-driven journalism** with its aim of helping us understand complex issues of social relevance and which integrates journalism with the new needs of contemporary informative understanding. The purpose of this paper is to discuss **data visualization** in **Brazilian journalism**, starting with what data visualization is and then, upon its concept and practical uses, determine how this view provides relevant advantages. By relevant advantages we mean matters of public interest with more critical, greater depth and context of content on **Big Data**. Initiatives that bring together images related to data and metadata occur on market practices, academic laboratories, as well as independent media. This **narrative system** is acted upon different human and non-human agents, whose structures are being built with machinic codifications, using databases that communicate with other layers until reaching a user interface. There is a need for new expertise from professionals, teamwork and basic knowledge, often in programming languages, statistics and operational tools to build dynamic narratives and increasingly involve the reader. It is important to think about content that is available to different formats. For this research we adopted a multi-methodological strategy and the assumptions of the centrality of communication that permeates all communication and informational activities across the board, whether analog or not. A view that requires resilience in the face of theoretical and methodological approaches, so that they are able to embrace and support the reflections for this dynamic field of study. To make propositions and interpretations, adopted based on the Database Digital Journalism paradigm, and the contributions of format concepts (RAMOS, 2012 and MACHADO, 2003), post-industrial journalism (COSTA, 2014), system narrative and antenarrative (BERTOCCHI, 2013) maturing as means of understanding the proposed object.

Keywords: data visualization, narrative system, Brazilian journalism, Big Data, data-driven journalism

Área de Concentração

Esta pesquisa insere-se na Área de Concentração denominada Teoria e Pesquisa em Comunicação, uma das áreas mantidas pelo Programa de Pós-Graduação em Ciências da Comunicação (PPGCOM) da ECA-USP. Esta área está voltada às reflexões epistemológicas, às definições teóricas e às propostas metodológicas para o estudo do fenômeno da comunicação em geral ou aplicada a modalidades específicas (comunicação interpessoal, grupal, massiva, tecnologicamente mediada e outras). Área que entende a comunicação como um campo de conhecimento interdisciplinar e transdisciplinar desde as suas origens em função da complexidade comunicacional e de suas repercussões das mais variadas ordens: histórica, estética, cultural, política, social e tecnológica.

Linha de pesquisa

O estudo encontra-se na linha de pesquisa do PPGCOM ECA-USP designada Comunicação e Ambiências em Redes Digitais. Essa linha mantém um foco nas reflexões epistemológicas e nos recortes teórico-metodológicos decorrentes da inserção do fenômeno da comunicação em ambiências de redes digitais sustentadas pelas tecnologias digitais de informação e comunicação (TICs).

Grupo de pesquisa

Esta pesquisa desenvolve-se também no âmbito do Com+, Grupo de Pesquisa em Comunicação, Jornalismo e Mídias Digitais. O grupo agrega pesquisas nas áreas de Políticas e Estratégias e Comunicação Digital e Tecnologias da Comunicação e Redes Interativas. Está vinculado ao Departamento de Jornalismo e Editoração da Escola de Comunicações e Artes da Universidade de São Paulo; sob coordenação da profa. Dra. Elizabeth Nicolau Saad Corrêa. Endereço eletrônico: <http://grupo-ecausp.com/>.

Financiamento

Esta pesquisa contou com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), agência do Ministério da Ciência, Tecnologia e Inovação (MCTI).

Recado aos leitores

As citações em língua estrangeira receberam livre tradução desta autora no corpo da dissertação. Os textos originais foram alocados nas notas de rodapé.

O tema deste trabalho nos forçou a utilizar algumas siglas, abreviações e jargões tecnológicos do mercado de Internet. Para facilitar a leitura criamos um índice com essas expressões.

Ao longo dos textos destacamos frases em negrito para reforçar os referenciais adotados na pesquisa.

Lista de Figuras

Figura I. 1 – Visualização de Dados, 2013.	41
Figura I. 2 – Nytimes.com, matéria Reshaping New York, 2013.	47
Figura I. 3 - Miamiherald.com, matéria Innocents Lost, vencedora do OJA 2014.	48
Figura I. 4 – Nytimes.com, matéria State Gun Laws Enacted in Year Since Newton, 2013.	49
Figura I. 5 – Esquema conceitual da visualização de dados nesta pesquisa	67
Figura II. 6 – Mapa de Snow, 1854.	70
Figura II. 7 – Globo.com, matéria A distribuição dos clientes e valores do "Swissleaks", 2015.	72
Figura II. 8 – Estadadados.com, projeto Basômetro, 2008.	77
Figura II. 9 – Encontro Hack d'Água' São Paulo, 2014	78
Figura II. 10 – Labic.net, cartografia A longa jornada do #MarcoCivil, 2012	80
<i>Figura III. 11 – O Sistema Narrativo no Jornalismo Digital (BERTOCCHI, 2013).</i>	107
Figura III. 12 – Versão do diagrama Chart Suggestions – A Thought-Starter, de Andrew Abela	116
Figura III. 13 – Matriz de Construção Inforgráfica (TEIXEIRA, 2014)	119
Figura IV. 14 – Familiarização com o tema visualização de dados.	129
Figura V. 15 – Estadadados.com, home do dia 14 de dezembro de 2015.	147
Figura V. 16 – Usabilidade da ferramenta Basômetro – taxa de governismo de candidato do PSC/PB	150
Figura V. 17 – Série de infográficos interativos explicam financiamento estudantil	155
Figura V. 18 – Gráfico interativo traz deputados com problemas na justiça.	163
Figura V. 19 – Ferramenta interativa aponta qual a chance de um projeto passar na Câmara	165
Figura V. 20 – Infográfico interativo permite comparar o desempenho das seleções da Copa	166
Figura V. 21 – A visualização de dados no jornalismo contemporâneo.	182
Figura 22 – Hashtag #7minutos1Denuncia – Twitter, 8 de março de 2016	188

Siglas e expressões

Para facilitar a leitura, segue relação dos significados de siglas e expressões presentes nesta dissertação.

Accountability – termo da língua inglesa que pode ser traduzido como responsabilidade com ética;

API (Application Programming Interface) – interface de programação de aplicativos;

App (Application) – aplicação, aplicativo ou software;

Backend – interface digital administrativa acessível por administradores do sistema;

Big Data – gigantesco armazenamento de dados tendo a velocidade, volume e variedade como pressupostos;

Business analytics – abordagem centrada em dados que combina a ciência de análise preditiva com capacidades avançadas de inteligência de negócios;

Cloud-computing (computação na nuvem) – refere-se à utilização da memória e das capacidades de armazenamento e cálculo de computadores e servidores compartilhados e interligados por meio da Internet;

CMS (Content Management System) – sistema de gerenciamento de conteúdos;

Copyleft ou creative commons – tipos de licença de direitos autorais que permitem a outras pessoas copiar, distribuir e executar a obra original;

Copyright – é um direito autoral a propriedade literária que concede ao autor de trabalhos originais direitos exclusivos de exploração de uma obra artística, literária ou científica, proibindo a reprodução por qualquer meio. É uma forma de direito intelectual;

Creative Commons – ontologia *web* para representação de licenças autorais;

CSS (Cascading Style Sheets) – linguagem de estilo que define a apresentação de documentos HTML ou XML;

Dadosfera – cenário contemporâneo diante da disposição de dados presentes na *web*, que modifica a forma de coleta e armazenamento de informações. Fator que justificaria a necessidade de curadores de notícias.

Data-driven journalism – jornalismo guiado por dados, jornalismo de dados, jornalismo em base de dados;

Data mining (mineração de dados) – funcionalidade que agrega e organiza dados, encontrando neles padrões, associações, mudanças e anomalias relevantes;

Data scrap ou data scraping (raspagem de dados) – técnica computacional na qual um programa extrai dados de saída legível somente para humanos, proveniente de um serviço ou aplicativo;

Dataentry – interface digital para entrada de dados;

Datastream – fluxo de dados;

Frontend – interface digital acessível por usuários finais;

GIS – Sistemas de Informação Geográfica;

Grounded Theory ou GT – método e metodologia de pesquisa;

Hipertexto – termo que remete a um texto ao qual se agregam outros conjuntos de informação na forma de blocos de textos, palavras, imagens ou sons, cujo acesso se dá através de referências específicas;

HTML (HyperText Markup Language) – linguagem de marcação para criação de sites *web*;

ICIJ (International Consortium of Investigative Journalists) – Consórcio Internacional de Jornalistas Investigativos;

Input / Output – entrada e saída de dados;

IP - Protocolo de Internet – identificação de um dispositivo (computador, impressora, em uma rede local ou pública);

Java – linguagem server-side como PHP e Python;

JavaScript – linguagem de programação interpretada;

JDBD - Jornalismo Digital de Base de Dados;

Machine Learning ou Aprendizado de Máquina – método de análise de dados que automatiza o desenvolvimento de modelos analíticos;

Metadados – dados sobre outros dados;

MJDBD – Modelo Jornalismo Digital de Base de Dados;

MOOC (Massive Open Online Course) – Curso Online Aberto e Massivo;

MySQL – sistema de gerenciamento de banco de dados que utiliza como interface a linguagem SQL, Linguagem de Consulta Estruturada, do inglês Structured Query Language;

Open data (dados abertos) – correspondem à ideia de que certos dados devem estar disponíveis para que todos usem e publiquem, sem restrições de direitos autorais e patentes;

Paywall – é um sistema de assinatura que tem sido utilizado por jornais e outros veículos de comunicação digitais que permite ao internauta o acesso a conteúdos restritos.

PC (Personal Computer) – computador pessoal;

PHP (Hypertext Preprocessor) – linguagem interpretada livre para criação de sites *web*;

PHP Admin – aplicativo *web* desenvolvido em PHP para administração do MySQL pela Internet;

Python – linguagem de programação de alto nível;

R – linguagem e também ambiente de desenvolvimento integrado para cálculos estatísticos e gráficos;

RAC – Reportagem Assistida por Computador;

Scrapping – no sentido deste estudo, raspagem de dados;

SEO (Search Engine Optimization) – otimização de sites para motores de busca;

SGC (Sistema de Gerenciamento de Conteúdos);

SKU (Stock Keeping Unit) – Unidade de Manutenção de Estoque está ligado à logística de armazenamento e designa os diferentes itens do estoque (código identificador);

Software – programa de computador;

Streaming – forma de transmissão instantânea de dados de áudio e vídeo através de redes.

Tagging – tagueamento, etiquetamento;

Template – modelo;

TI – Tecnologia da Informação;

TICs – Tecnologias da Informação e Comunicação;

Timeline (linha do tempo) – ordem das publicações feitas nas plataformas sociais online;

Web ou WWW – sistema de documentos em hipermídia interligados e executados na Internet

WS – *Web* Semântica – geração de conhecimento pela máquina;

Yottabyte – unidade de medida que equivale a 10^{24} de dados.

SUMÁRIO

INTRODUÇÃO	17
Objeto e justificativa	24
Quadro teórico de referência	28
Objetivos	31
<i>Objetivo geral:</i>	31
<i>Objetivo específico:</i>	31
<i>Objetivo teórico:</i>	32
<i>Objetivo prático:</i>	32
Mapa da dissertação	32
I. VISUALIZAÇÃO DE DADOS NO JORNALISMO BIG DATA	35
Dado e informação	35
A visualização de dados no jornalismo contemporâneo	41
<i>Exemplificando</i>	46
O jornalismo na era Big Data	50
Jornalismo Digital de Base de Dados (JDBD) e o jornalismo de dados	56
Jornalismo de dados: entre o senso comum e a ciência	62
Considerações	66
II. A ENTREVISTA AOS DADOS	69
A avalanche de informações	69
<i>O passo-à-passo</i>	73
Visualização de dados no jornalismo digital brasileiro	76
Características inerentes	81
<i>Equipes</i>	81
<i>Tempo de produção</i>	82
<i>Fontes dos dados</i>	82
Experiência narrativa e usabilidade	87
Cenário político e discussões sobre a Internet	89
<i>Liberdade e privacidade na rede</i>	91
Considerações	96
III. O FAZER JORNALISMO DE DADOS E A VISUALIZAÇÃO NO SISTEMA NARRATIVO	98
Diferentes olhares e novas práticas	98
<i>Princípios da narrativa contemporânea</i>	102
Jornalismo como sistema: seleção de informações e formato narrativo	105
<i>Escolhas e caminhos da modelagem jornalística</i>	111
<i>Design</i>	114
Valor na contemporaneidade	122
Considerações	123
IV. CONSTRUINDO A PARTIR DA DESCONSTRUÇÃO – CONSIDERAÇÕES METODOLÓGICAS	126
Diferentes olhares	126
Grounded Theory ou Teoria Fundamentada	136
<i>Características</i>	136
<i>Codificação e análise</i>	137
<i>Avaliando uma Grounded Theory</i>	138
Aplicando a Grounded Theory no blog Estadão Dados	140
<i>Acompanhamento do blog</i>	140
<i>Entrevistas semi-estruturadas</i>	140

<i>Caminhos de estruturação.....</i>	<i>141</i>
Considerações.....	143
V. BLOG ESTADÃO DADOS.....	145
Lindando com os dados.....	145
Blog Estadão Dados.....	147
Basômetro	149
<i>Os desdobramentos do fazer.....</i>	<i>151</i>
Farra do Fies.....	153
<i>Os desdobramentos do fazer.....</i>	<i>158</i>
Reflexões desencadeadas	160
<i>Pontos jornalísticos destacados</i>	<i>161</i>
<i>Finalidade do conteúdo.....</i>	<i>161</i>
<i>Testes, adaptações e autodidatismo.....</i>	<i>163</i>
<i>Faro jornalístico e formação das equipes.....</i>	<i>166</i>
<i>Desafios do novo.....</i>	<i>169</i>
<i>Previsão.....</i>	<i>170</i>
<i>Repercussão do conteúdo.....</i>	<i>173</i>
<i>Armazenamento e recuperação de dados (datalização)</i>	<i>174</i>
<i>Solução e validação das peças</i>	<i>177</i>
<i>A lógica da produção dos conteúdos.....</i>	<i>179</i>
Sustentação teórica vinculada a core category	181
Considerações.....	184
VI. CONSIDERAÇÕES FINAIS.....	187
REFERÊNCIAS	196

INTRODUÇÃO

Esta dissertação é um estudo sobre a **visualização de dados no jornalismo** em um cenário de grande volume e velocidade de tráfego de dados na Internet. Apesar de apontar resultados finais como gráficos ou infográficos, a proposta não é abordar tais imagens no jornalismo, mas sim discutir sobre os processos e esclarecer conceitos da visualização de dados no jornalismo. Vivenciamos um momento onde há a fusão de informações e relações nos diferentes ambientes proporcionados pela *web*, onde muitos deles se complementam, ganhando dinamismo e interação com a Internet. Ao tratar da visualização de dados neste pesquisa podemos dizer que é um determinado tipo de visualização em um determinado modo de fazer jornalismo, em um cenário que se convencionou chamar de *Big Data*¹, fenômeno que envolve o armazenamento, velocidade, volume e variedade de dados disponíveis hoje na *web*.

A visualização de dados está cada vez mais presente nos meios de comunicação e iniciativas jornalísticas que buscam conteúdos relevantes para a sociedade e inovadores na forma de apresentação da informação para os leitores. Ao visitar na *web* as páginas de jornais nacionais e internacionais, é possível perceber o crescimento das informações visuais com seções exclusivas de dados, assim como as já presentes sobre infografias². As iniciativas que reúnem imagens relacionadas a um grande volume de dados ocorrem nas práticas de mercado, laboratórios acadêmicos³, assim como em mídias independentes⁴, como observaremos ao longo da pesquisa. As diferentes empresas de mídia tem se aperfeiçoado e disponibilizado na rede conteúdos aprofundados sobre situações contemporâneas. O jornalismo de dados preenche as lacunas de conteúdos mais densos e contextualizados aos anseios de uma sociedade

¹ O termo *Big Data* tem origem na tecnologia da informação (TI). É focado no gigantesco armazenamento de dados tendo a velocidade, volume e variedade como pressupostos. Algumas empresas de TI, como a Oracle, ainda consideram o valor como predeterminante. Já a empresa de consultoria SAS considera outros dois pontos, a variabilidade e complexidade como pressupostos da *Big Data*. Retomaremos este assunto ao longo da dissertação.

² Entre os diversos exemplos, nacionalmente podemos destacar os blogs dos jornais Estado de S. Paulo (<http://blog.estadaodados.com>) e Folha de S. Paulo (<http://folhaspdados.blogfolha.uol.com.br>) e internacionalmente os jornais The Guardian (<http://www.theguardian.com/data>) e The New York Times (<http://www.nytimes.com>).

³ Entre as diversas iniciativas, o laboratório de estudos sobre Internet e Cultura do Departamento de Comunicação da Universidade Federal do Espírito Santo (LABIC/UFES), disponível em: <http://www.labic.net>. Um segundo seria o laboratório em mídias e métodos digitais da Universidade Federal do Rio de Janeiro MídiaLab, disponível em: <http://medialabufrrj.net/tag/visualizacao-de-dados/>

⁴ Um exemplo é o Hacks Hackers SP, que visa unir pessoas em ações colaborativas que envolve explorar tecnologias para filtrar e visualizar informações. Disponível em: <http://www.meetup.com/Hacks-Hackers-Sao-Paulo/>. Outro é o projeto InfoAmazonia, que agrega dados e notícias sobre a Amazônia. Ele é sustentado por uma rede de organizações e jornalistas que oferecem atualizações constantes dos nove países da região. Disponível em: <http://infoamazonia.org>.

interconectada, que cada vez mais se representa a partir de imagens e números.

O intuito desta dissertação é problematizar a visualização de dados no jornalismo brasileiro partindo exatamente do que é a visualização de dados, mais especificamente no jornalismo de dados⁵, e diante dos apontamentos sobre seu conceito e prática questionar como esta visualização proporciona diferenciais relevantes no jornalismo. Por relevantes entendemos pautas de interesse público, que envolvem maior criticidade, maior aprofundamento e contextualização dos conteúdos no *Big Data*. Propõe-se o estudo da visualização de dados no momento de um jornalismo pós-industrial⁶, onde ocorrem significativas mudanças no modo de “jornalisticar”⁷. Há a proximidade dos jornalistas com a audiência, em relações não mais hierárquicas. Há a possibilidade de se aprofundar os fatos diante do anseio de uma sociedade participativa e cada vez mais questionadora. Há a necessidade de novas expertises por parte dos profissionais de mídia como o conhecimento básico, muitas vezes, em linguagem de programação, estatística e operacionalização de ferramentas na construção de narrativas dinâmicas e que cada vez mais envolvam o leitor. Sendo importante o pensar sobre um conteúdo que seja disponível para diferentes formatos. Mesmo não sendo especificamente o objetivo deste estudo, o que automaticamente nos leva a rever a presença da universidade na formação dos aprofundamentos diante de grande volume de dados e correlações entre dados que possibilitam novas interpretações de fenômenos.

A disponibilidade de recursos tecnológicos promove uma revolução na capacidade do indivíduo em produzir e circular informação e em reunir estes dados de forma que gerem novos sentidos. Neste cenário, entender como alguns processos se estabelecem são fundamentais para os estudos da Comunicação, referendando esta área como uma das mais importantes para a compreensão da sociedade, suas crenças e manifestações. Importante para situar e questionar o jornalismo enquanto prática social (mediação, credibilidade, agendamento, legitimidade e fluxo noticioso) que se reconfigura na

⁵ As variações do termo jornalismo de dados, *data-driven journalism*, jornalismo de base de dados e jornalismo guiado por dados serão amplamente utilizadas, tendo o mesmo significado conceitual, que será melhor explicado ao longo da dissertação.

⁶ (...) jornalismo já não mais organizado consoante uma lógica industrial em cascata produtiva; mas com marcas mais complexas: mais atores atuantes, mais circularidade, mais algoritmos, mais inteligência artificial, mais computação em seu interior. (BERTOCCHI, 2013, p. 10)

⁷ A professora do departamento de jornalismo e editoração da Escola de Comunicação e Artes da Universidade de São Paulo, Dra. Elizabeth Saad, aborda o termo na mesa 2 #narrativas do evento “Quem Mexeu No Meu Jornalismo?” realizado pela Faculdade Cásper Líbero. Disponível em: <http://casperlibero.edu.br/assista-ao-vivo-quem-mexeu-no-meu-jornalismo/>. Acesso em: 08 de maio de 2014.

nova ecologia midiática. Um jornalismo inserido em uma sociedade líquida (BAUMAN, 2011)⁸, que sinaliza o término de um dado modelo de jornalismo informativo diário, que passa por um período de instabilidades, num contínuo de mudanças que aparentemente desencadeiam processos de alargamento das fronteiras do campo comunicacional.

Os dados na *web* tornam-se elementos quantitativos e manipuláveis quando transcodificados para a linguagem computacional. Uma vez transformados em linguagem numérica possibilitam diferentes formas de visualização a partir de ferramentas da computação que agilizam o trabalho de fazer correlações sejam de dados estruturados ou não⁹. A prática da visualização de dados é cada vez mais presente também em diversos setores, sejam eles públicos ou privados. A coleta e a análise correta dos dados tem gerado impactos expressivos em áreas como finanças, saúde e educação¹⁰. Onde epidemias podem ser previstas com maior precisão; dados gerados na educação podem ser utilizados para prever as demandas e as problemáticas do setor, e, como ocorreu as vésperas das eleições presidenciais norte-americanas, em 2012, prever o resultado dos colégios eleitorais com grande precisão¹¹.

Hoje falamos em *yottabytes*¹² de dados trafegados nos Protocolos de Internet (IPs). Os dados estão disponíveis em diferentes extensões, dentro e fora da *web*, oriundos de diversas áreas, sejam coletas estatísticas, pesquisas científicas, cálculos matemáticos, transações financeiras. Em uma sociedade que se estrutura de forma complexa, os desafios adquirem distintas abordagens para que as reflexões progridam com a mesma velocidade com que o mundo se modifica, principalmente num cenário que se reconfigura diante da comunicação generalizada (VATTIMO, 1992). Tem-se a dificuldade de discutir o presente, a dinâmica social contemporânea com

⁸ BAUMAN, Zygmunt. 44 cartas do mundo líquido moderno. Tradução: Vera Pereira. Rio de Janeiro: Jorge Zahar, 2011.

⁹ Nem todas as informações geradas e armazenadas por computador se encontram em bancos de dados, em linhas e tabelas estruturadas. A maior parte não está estruturada, são informações guardadas em documentos, e-mails, vídeos, conteúdos em redes sociais digitais, estruturados nas mais diferentes formas e com suas complexidades. TAKAI, O.K. Takai; ITALIANO, I.C; FERREIRA, J.E. Introdução a Banco de Dados. DCC-IME-USP, 2005. Disponível em: <http://www.ime.usp.br/~jef/apostila.pdf>. Acesso em 5 de agosto de 2013.

¹⁰ De acordo com o relatório *Big data, big impact: new possibilities for international development*¹⁰, produzido pelo Fórum Social Mundial realizado em 2012, em Davos, na Suíça. Disponível em: <http://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development>.

¹¹ Outras informações podem ser obtidas na página do Observatório da Imprensa. Disponível em: http://www.observatoriodaimprensa.com.br/news/view/_ed720_estatistico_do_nytimes_previu_resultado_dos_colegios_eleitorais. Acesso em: 13 de julho de 2013.

¹² De acordo com o *BigData-Startups*, 1 yottabyte (10^{24}) equivale a 250 trilhões de DVDs de dados, mas já se falam nos BrontoBytes de dados (10^{27}) no universo digital.

protagonistas vivos, atuantes e sem parâmetros para as situações que se apresentam cotidianamente. Conceitualmente partimos de alguns pressupostos, entre eles:

1º) de que a visualização de dados se distingue de outras práticas visuais por envolver a geração de imagens de forma automatizada, sendo necessária para a sua realização desenhos algorítmicos¹³ que resultam em riqueza no detalhamento de dados, principalmente porque eles podem ser trabalhados em grande escala. Inicialmente, a principal preocupação na sua concepção não está na estética da apresentação das informações, mas sim na melhor forma de evidenciar fatores que muitas vezes estão ocultos nas diferentes extensões e correlações dos dados (ESPERANÇA, 2014). Por isso, muitas vezes, a sua leitura é feita por especialistas e destinada a um pequeno público.

2º) no jornalismo, as inúmeras possibilidades de elementos combinatórios, gráficos estatísticos, interativos, aliados ou não a uma narrativa fazem parte do jornalismo de dados (*data-driven journalism*, jornalismo guiado por dados ou jornalismo em base de dados), inserido no paradigma Jornalismo Digital de Base de Dados (JDBD)¹⁴. O *data-driven journalism* representa uma forma diferenciada de jornalismo, onde o seu objetivo é encontrar as pérolas da notícia (CRUCIANELLI, 2014), contando para isso com ferramentas digitais para manusear os dados. Reúne as técnicas do jornalismo investigativo, de profundidade, de precisão, analítico e da reportagem assistida por computador¹⁵. Para a pesquisadora Suzana Barbosa (2007), o jornalismo de dados traz as características da quarta fase do jornalismo digital¹⁶, uma transcodificação do jornalismo em um novo formato, o das bases de dados, com funções de indexar objetos multimídia (sons, imagens, gráficos), armazenar material produzido e de

¹³ “Um algoritmo é uma sequência finita de instruções bem definidas e não ambíguas, cada uma das quais pode ser executada mecanicamente em um período de tempo finito e com uma quantidade de esforço finita”. Fonte: <http://pt.wikipedia.org/wiki/Algoritmo>. Acesso em: 10 de fevereiro de 2015. ¹⁴ Na comunicação digital, o algoritmo tem a missão de expurgar informações indesejáveis, oferecendo apenas o que o usuário julgaria eventualmente o mais relevante para si, conforme um modelo de negócio definido ou de acesso às informações também previamente determinado pelo proprietário do algoritmo. Já a *web* semântica é um espaço digital onde máquinas conseguem entender o significado dos dados em diversos contextos e sem interpretações sobrepostas. CORRÊA, Elizabeth S. BERTOCCHI, Daniela. A cena cibercultural do jornalismo contemporâneo: *web* semântica, algoritmos, aplicativos e curadoria, 2012. Disponível em: <http://www.matrizes.usp.br/index.php/matrizes/article/view/340> Acesso em: 5 de junho de 2013.

¹⁴ Para Suzana Barbosa (2007) o *data-driven journalism* é uma consequência do desenvolvimento alcançado pelo jornalismo digital, pela tecnologia de bases de dados, pela expansão da Internet e de iniciativas de informações livres ("*open data*"), e pelo barateamento dos processos produtivos online.

¹⁵ Estas práticas jornalísticas e suas características, mesmo que sutis, serão melhor discutidas no primeiro capítulo. De antemão temos o jornalismo investigativo como uma abordagem que se trata de um tema de relevância social que alguém quer manter em segredo. O jornalismo de profundidade é semelhante ao jornalismo investigativo, com a diferença que ele não é mantido em segredo. O jornalismo de precisão envolve o uso de métodos da investigação social aplicados ao jornalismo em investigações qualitativas e quantitativas. Como o próprio nome diz, o jornalismo analítico faz análises com o intuito de oferecer provas baseadas em interpretações de dados. Já o jornalismo assistido por computador é o processo que utiliza a assistência de computadores durante a relação e processamento de dados.

¹⁶ A quinta fase tem as mídias móveis como agentes de um novo ciclo de inovação, a existência de um nível expressivo de replicação de conteúdos na distribuição multiplataforma/cross-media (BARBOSA, 2013).

arquivo (memória), agilizar produções, compor conteúdos para a *web*, recuperar, informações e, principalmente, cruzar dados que gerem uma nova informação visual e dinâmica.

Partimos da ideia de que mesmo com a presença de ferramentas gratuitas, tutoriais de uso, é difícil produzir um fluxo constante de jornalismo de dados de qualidade de forma barata e com uma equipe reduzida (concordando com CAIRO, 2014). Demandam-se diferentes competências. As técnicas utilizadas não são novas, são métodos quantitativos, análises, o que torna esta prática mais uma evolução, do que uma revolução ao se trabalhar com dados. Os leitores são participativos, muitas vezes podem conhecer profundamente o assunto, seja porque as bases de dados são públicas ou por serem especialistas na temática. Os profissionais que atuam nas equipes de dados tem que ter uma compreensão básica de diferentes disciplinas, entender a origem dos dados, saber como eles estavam reunidos, como serão filtrados e processados. Principalmente, precisam contextualizar os fenômenos e fazer as perguntas certas aos dados (CRUCIANELLI, 2014). “As bases de dados não contam histórias, não têm começo e nem fim, e estão potencialmente sendo continuamente alimentadas, portanto, em constante mutação”. (RAMOS, 2011, p. 23).

3º) Em concordância com a pesquisadora Daniela Bertocchi (2013), acreditamos que narrativa digital jornalística é sistema, processo, fluxo e que dentro do sistema narrativo a visualização de dados “revela-se como o aproveitamento do agenciamento entre as camadas de dados e metadados (...) o qual privilegia a apresentação visual relevante e de impacto. (BERTOCCHI, 2013, p. 180).

Ao colocar a narrativa como sistema, decidimos aqui caminhar rumo às camadas mais subterrâneas de sua modelagem: descemos primeiramente à antenarrativa (a narrativa em potencial) para, em seguida, voltarmos em direção à superfície, observando o formato no qual ela se manifesta (com quais tipografias, cores, formas, *design*, gêneros textuais), expandindo nosso olhar até a camada do usuário (lugar em que ocorrem enunciações e manipulações externas à narrativa e que a modificam, ou seja, o entorno do sistema narrativo). (BERTOCCHI, 2013, p. 40).

Além destes pressupostos, adotamos como primordial a contextualização do momento vivenciado pela sociedade por estarmos cada vez mais presentes na rede digital, com

fluxo informativo cotidiano e com as inúmeras trocas na *web*, nas quais deixamos rastros passíveis de serem coletados e transformados em informação. O pensador Manuel Castells (2003) aponta que em relação aos avanços tecnológicos, a Internet se caracteriza mais do que como uma tecnologia e sim como um meio de comunicação, de interação e de organização social. Neste cenário é um desafio trabalhar com dados no jornalismo, com narrativas inovadoras e com um público cada vez mais participativo. Para o estudioso Carlos Scolari (2008), nenhuma das teorias da Comunicação do século XX é capaz de explicar as questões atuais do meio. A Comunicação é um objeto de múltiplas dimensões, que deve ser analisada de diversos pontos de vista. Neste sentido, as zonas mais inovadoras de conhecimentos são as fronteiriças, os espaços de tradução. Para Scolari estamos diante de uma ruptura de categorias que fundaram o processo cultural, do consumo e da produção comunicacional. Os recursos tecnológicos possibilitam um fazer jornalístico diferente, novas formas de interação com o leitor em um tempo de produção, execução e consumo múltiplos. Há um maior trabalho, por parte do jornalista, diante do leque de opções em apresentar o conteúdo, de contextualizá-lo nos diferentes formatos.

Chegar a este estudo envolve uma trajetória de curiosidades sobre o tema e vivências práticas de partes do processo dentro de veículos de grande mídia (seja integrando equipe de implantação de portal de notícia¹⁷ ou atuando na redação e alimentação de blogs ou redes sociais¹⁸). Na graduação o trabalho de conclusão do curso de Comunicação já abordava a temática¹⁹ do jornalismo digital²⁰. No entanto, o *start* para estudar com mais profundidade o jornalismo de dados e a visualização de dados ocorreu após a experiência de coordenação de um projeto de responsabilidade social que envolvia o mapeamento de reclamações de ouvintes sobre a infraestrutura dos bairros em que moravam. A equipe de reportagem do projeto visitava diariamente os bairros da capital e grande São Paulo para ouvir as demandas dos moradores. As reclamações eram catalogadas e encaminhadas aos órgãos públicos e existia todo um processo de veiculação das respostas na emissora e no blog. Após quatro anos de atividade nas ruas, surgiu a ideia de mensurar as ações, quantificá-las principalmente

¹⁷ Implantação do portal Imirante.com, do Sistema Mirante de Comunicação, São Luís/MA, 2003.

¹⁸ Entre 2008 e 2012, Rádio Globo SP, <http://radioglobo.globo.com/home/HOME.htm>.

¹⁹ Jornalismo no suporte on-line: uma análise de caso dos sites maranhenses, 2005, Universidade Federal do Maranhão.

²⁰ Evitaremos fazer distinções entre jornalismo digital e não-digital, pois para nós essa dicotomia não se sustenta mais. Na qualificação desta pesquisa, em 2015, um dos pontos argumentados pela pesquisadora Daniela Bertocchi foi da não distinção entre online e off-line já que hoje vivenciamos um tudo junto e misturado. Concordamos com este pensamento, mas não nos aprofundaremos nele para não fugir da temática da pesquisa.

por três fatores: primeiro para saber exatamente com o que estávamos lidando em relação a dimensão e impacto da ação. Em segundo para justificar a presença do projeto no cenário de crise vivido pelo rádio AM em São Paulo, onde os resultados da ação poderiam ser apresentados para o departamento comercial (tentativa de aquisição de patrocínio) e para os acionistas (manutenção da iniciativa por parte da empresa). Um terceiro fator é que o banco de dados próprio serviria para classificar as principais reclamações de cada bairro, região, cidade, quais obtiveram respostas de órgãos públicos, entre inúmeras outras correlações que resultariam em uma apropriação e projeção dos problemas vivenciados cotidianamente pelos ouvintes. A ideia de trabalhar com os dados e construção de uma base de dados própria no projeto entrou em prática a partir do esforço de duas jornalistas que não tinham conhecimento algum em arquivamento, catalogação, ferramentas ou demandas do gênero, assim como faltava habilidade em qualquer forma de resgate deste banco. Foi um tremendo esforço (próprio e manual) que resultou no estímulo a retomar os estudos especificamente nesta área diante do atual cenário de gerenciamento de dados, da presença de ferramentas gratuitas, além da enorme vontade de estar em sala de aula.

Este estudo tem como proposta uma reflexão e a audácia de tentar ir além, rumo aos questionamentos que contribuam para uma produtividade jornalística com mais rigor e cientificismo. Tendo a visão de Comunicação como um campo de conhecimento interdisciplinar e não apenas como um aspecto da sociedade e ao propor o estudo de um fenômeno emergente aplicado aos usos mediados da tecnologia, o presente estudo faz parte da Área I – Teoria e Pesquisa em Comunicação, do programa Ciências da Comunicação, da Escola de Comunicação e Artes, da Universidade de São Paulo. Integrado a linha III – Comunicação e Ambiências em Redes Digitais – que se ocupa da reflexão sobre complexidades decorrentes dos impactos das Tecnologias da Informação e Comunicação, da construção de sentido nas linguagens digitais.

Para o desenvolvimento da pesquisa foi adotada uma estratégia multimetodológica. Como percurso natural do início do mestrado, recorreremos as referências bibliográficas que nos atualizassem sobre o cenário do jornalismo contemporâneo e o Big Data. Com a pesquisa empírica percebemos a dificuldade de tentar olhar e compreender o objeto partindo de teorias e metodologias já consolidadas. No entanto, permanecia o desafio de produzir uma pesquisa com um rigor empírico e com embasamento teórico

condizente para entender os fenômenos jornalísticos. Por conta disso, após a qualificação, o olhar foi invertido para os dados e o que eles poderiam dizer. Uma maneira de realizar as análises dos dados qualitativos, de modo que estas fossem enraizadas nos dados, trouxe a Grounded Theory²¹ (GT) ou Teoria Fundamentada para esta pesquisa. O objetivo foi construir categorias analíticas a partir dos dados de pesquisa e, por conseguinte, respeitar o fenômeno seguindo indicações oriundas destes dados. Como um estudo iniciante no mestrado, na Grounded Theory e na visualização de dados, partimos dos pressupostos da GT e sempre do começo: o que são, para daí entender o fenômeno. Como observação aprofundada o foco foi o jornalismo de dados e a prática da visualização no blog Estadão Dados, do jornal O Estado de S. Paulo.

Objeto e justificativa

O que é a visualização de dados no jornalismo? Como ela se insere na prática e na narrativa do jornalismo de dados proporcionando diferenciais relevantes? O objeto deste estudo é o jornalismo de dados no Brasil, com estudo de caso do blog Estadão Dados, do jornal O Estado de S. Paulo. A escolha se deu por ter no Blog²² tanto conteúdos gráficos, quanto projetos em visualização, pelas constantes atualizações²³ e por estar em um cenário brasileiro de investimentos nesta área. É um dos poucos em funcionamento contínuo e não somente em projetos especiais e esporádicos, o que nos possibilitou um estudo com maior embasamento de campo diante dos números de experiências em prática. Buscamos com o estudo de caso, um método de coletar dados do corpus, a possibilidade de aprofundar a pesquisa, de avaliar quantitativamente e qualitativamente os processos envolvidos na produção de conteúdos que envolvam o manuseio de grande volume de dados. O estudo de caso permitiu a profundidade necessária para obter parte da compreensão desejada. A pesquisa foi prosseguida pela Grounded Theory, que fez emergir processos subjacentes às afirmações dos participantes e ao fenômeno observado. O capítulo IV é

²¹ No Brasil, a GT é também chamada de Teoria Fundamentada nos Dados, de Teoria Fundamentada em Dados, ou ainda, de Teoria Fundada. Adotamos neste estudo a terminologia Grounded Theory (GT), em inglês, como no original da criação do método, porque o termo Teoria Fundamentada (seja “em Dados” ou “nos Dados”) é tradução de uma das etapas da GT, a última etapa na qual encontramos o resultado do processo.

²² Como meio de distinção, toda vez que aparecer a palavra Blog, com letra inicial maiúscula, estamos nos referindo ao blog Estadão Dados.

²³ O acompanhamento ao Blog acontece desde 2013.

destinado aos esclarecimentos metodológicos. Fator essencial e fundamental para justificar as escolhas adotadas ao longo do caminho. Por mais contraditórias que possam ser em um primeiro momento, encontramos meios coesos de torná-las sustentáveis.

No jornalismo, a visualização de dados é usada para análises de tendências, no cruzamento de dados e como busca de informações no jornalismo investigativo (RAMOS, 2011). Informações que em decorrência do volume, variedade e diferentes extensões dos dados só podem ser agrupados e posteriormente correlacionados a partir de mediação tecnológica. Inicialmente, o teórico russo Lev Manovich (2006) apontou que o computador não delimitava o jornalismo digital, mas sim a condição digital o delimitava. Junto com cada meio, novos signos eram gerados e muitas vezes unindo diferentes códigos: gestual, verbal, sonoro, visual (estático ou em movimento). Para ele, em um primeiro momento, a forma como os novos meios, em geral, eram criados, distribuídos, guardados e arquivados em computadores definiam a lógica e influenciavam de maneira significativa na camada cultural e também nos próprios meios de comunicação.

Podemos pensar os novos meios de comunicação, em geral, constituídos por duas camadas distintas: a “camada cultural” e a “camada de informação” ou “informática”. Entre os exemplos de categorias que pertencem à camada cultural, temos a enciclopédia, a história, a história e o enredo, a composição e ponto de vista, mime-se e catarse, a comédia e a tragédia. Enquanto exemplos de categorias da camada de informação temos os processos e o armazenamento de dados (como pacotes de dados que são transmitidos através da rede), triagem e harmonização, função e variável, a linguagem de computador e estrutura de dados (...). (MANOVICH, 2006, p. 93).

Ou seja, para Manovich (2006), a camada informática afetava a camada cultural e o mundo passava a ser modelado pela forma como os dados eram representados. No entanto, Manovich (2013) revisou o seu conceito de base de dados afirmando que o atual cenário da *web* (que sofreu o impacto de novas forças econômicas, sociais e tecnológicas (*web* comercial, redes sociais, computação mobile) conduz cada vez mais ao fluxo de dados, ao chamado *data stream*:

No meu artigo de 1998 chamei esta forma informativa de banco de dados, em oposição à organização informativa dominante: a narrativa. Usei a

palavra “banco de dados” ou “base de dados” para descrever um catálogo de objetos que não tem uma ordem padronizada de classificação. (Metaforicamente podemos dizer que num catálogo os objetos são organizadas no espaço em vez do tempo). (...) “Quero sugerir que, em mídia social, tal como se desenvolveu até agora (2004-2012), banco de dados não são mais regras. Em vez disso, a mídia social apresenta uma nova forma: um fluxo de dados. Em vez de navegar ou pesquisar uma coleção de objetos, o usuário experimenta o fluxo contínuo de eventos”. (MANOVICH, 2012, tradução nossa)²⁴.

O banco, a base de dados, continua com as informações-chave, mas coexistindo com outras formas, entre elas a visualização de dados. Para Ferdinand de Saussure (2006), o ponto de vista cria o objeto, ou seja, só existe a partir do significado que damos a ele. Boaventura de Sousa Santos (1989) traz a noção de que a ciência é um instrumento legitimador de escolha dos objetos de estudo, bem como dos problemas a serem resolvidos na realidade, onde o papel da comunidade científica, além de presumir os focos de ação, é também o de mediação entre o conhecimento científico e a sociedade. Acompanhar, à luz da ciência, como o jornalismo está construindo suas narrativas, captando e apresentando as questões sociais latentes utilizando e correlacionando o enorme volume de dados presentes no ciberespaço, no cenário digital, mostra-se relevante não só para profissionais da área, estudantes, mas para os próprios leitores, que são o público alvo do jornalismo.

Em toda pesquisa, mas especialmente na área em que o presente projeto se insere, Teoria e Pesquisa em Comunicação, deve-se mostrar as explorações teóricas e metodológicas compreendendo como ela pode contribuir para trabalhos futuros e para o avanço de uma linha de pesquisa. No segundo capítulo do livro *Crítica metodológica, investigação social e enquete operária*, Michel Thiollent (1980) traz as concepções do filósofo francês, Pierre Bourdieu, entre elas a questão da neutralidade em Ciências Sociais. Um dos primeiros critérios epistemológicos ao qual se refere Bourdieu consiste em considerar que toda técnica é uma teoria em atos, onde não existe uma coleta de dados sem pressupostos teóricos e é preciso estar preparado, ter um instrumental prático e teórico na hora de investigar.

²⁴ “In my 1998 article *Database as a Symbolic Form* I called this information form a “database” and opposed it to the historically dominant way of organization information – a narrative. I used the word “database” to describe a catalog of objects that does not have a default sort order. (Metaphorically, we can say that in a catalog the objects are organized in space rather than in time.). (...) “I want to suggest that in social media, as it developed until now (2004-2012), database no longer rules. Instead, social media brings forward a new form: a data stream. Instead of browsing or searching a collection of objects, a user experiences the continuous flow of events” (MANOVICH, 2012). Disponível em: <<http://lab.softwarestudies.com/2012/10/data-stream-database-timeline-new.html>> Acesso em: 10 de junho de 2014.

A mensuração, os instrumentos quantitativos e, de modo geral, todas as operações da prática sociológica, desde a elaboração de questionários e a codificação até a análise estatística, são teorias em atos a título de procedimentos de construção, consciente ou não, dos fatos e de relações entre os fatos. (BOURDIEU, 1973, p. 59)

Para Thiollent (1980), as técnicas mais neutras, ao serem incorporadas à pesquisa sociológica, acabam por funcionar como teorias particulares sobre a representação do objeto investigado.

Cada técnica contém instrumentos particulares cujo uso envolve pressupostos teóricos. Muitos epistemólogos concordam com este ponto de vista. G. Bachelard, por exemplo, sublinhou: "Na ciência moderna, um instrumento é verdadeiramente um teorema reificado" (BACHELARD, 1975, P. 140). Podemos considerar que os "teoremas reificados" ou os elementos de "teoria em atos" incluídos na pesquisa pelo uso das técnicas devem permanecer sob vigilância do pesquisador para evitar sua autonomização para com a teoria propriamente sociológica e, também, para superar o mito do objetivismo observacional (THIOLLENT, 1980, p. 44).

Habita-se atualmente uma paisagem onde a digitalização modifica a maneira como as pessoas se relacionam com outras pessoas, com produtos e com conteúdos. Neste cenário, empresas buscam mapear o seu público consumidor; instituições acadêmicas visam conhecer os atores sociais e seus comportamentos a partir do digital; governos em todo o mundo tendem a aumentar a transparência, disponibilizando dados referentes à administração pública sob licenças abertas (*open data*); as mudanças são inúmeras e envolvem grandes volumes de dados disponíveis na *web*. As formas de acesso, visualização e manipulação desses dados têm aumentado à medida que novos meios e ferramentas são descobertos. Assim como na sociedade, a visualização de dados no jornalismo está em processo evolutivo, introduzindo esta prática dentro das iniciativas das organizações jornalísticas, com adoção de novas ferramentas, rotinas produtivas, formatos e modos diferenciados de apresentação da informação.

O campo Comunicacional, especialmente o jornalístico, sofre o impacto de diversas forças como: do mercado (negócios digitais e processos lucrativos); da audiência (interação e usuários também produtores de conteúdo) e da academia (qualificação do profissional). Há outras competências que devem ser adquiridas, entre elas saber trabalhar em equipe. Os trabalhos unem diversos recursos como áudio, vídeo, texto,

em narrativas, muitas vezes, não-lineares. Conhecer o leitor torna-se ferramenta importante diante da facilidade de dispersão e de perda de audiência. O jornalismo pós-industrial exige um trabalho maior em recursos, equipes, investimentos na produção de um conteúdo com mais clareza e precisão para os diferentes públicos. E entender como as práticas acontecem neste cenário de velocidade, de participações múltiplas do público, compartilhamentos e grande volume de dados e suas representações em imagens e ferramentas interativas mostra-se relevante para profissionais e estudantes da área, para leitores interessados no jornalismo contemporâneo e profissionais de áreas correlatas como design, arquitetura de informações, computação, estatística, entre outros.

Quadro teórico de referência

A Comunicação não possui um único paradigma ou é um campo limitado a apenas uma disciplina. A especificidade do campo está no atravessamento dos saberes de áreas como Sociologia, História, Filosofia, Linguística, Economia, Psicologia, Antropologia, Computação, entre outros. Para Klaus Jensen (2008), a pesquisa em Comunicação é inter, trans, multidisciplinar e tradicional. Na visão de Pierre Bourdieu (1983), o campo científico tem regras próprias onde outras disciplinas são re-significadas quando adentram o campo da Comunicação. Para Edgar Morin (2006), é preciso ter em mente a noção da complexidade ao combinar diferentes perspectivas sob um novo ponto de vista, no caso o da Comunicação. É preciso que o conhecimento faça sentido no conjunto, situar informações e dados em seu contexto para que adquiram sentido. A emergência da visualização de grandes volumes de dados tornou-se uma necessidade crescente na sociedade e no jornalismo, e com isso, uma forma de extrair e analisar a informação em profundidade. Por ser um tema complexo da contemporaneidade, foram muitas idas e vindas para cercar o objeto conceitualmente. A teoria visa trazer elementos para se entender o que é a prática, mas os dados, neste estudo, trouxeram muitos outros elementos. Para se fazer as proposições e interpretações adotou-se como base o paradigma Jornalismo Digital em Base de Dados, tendo as contribuições dos conceitos de formato (RAMOS, 2012 e MACHADO, 2003), de jornalismo pós-industrial (COSTA, 2014), sistema narrativo

e antenarrativa (BERTOCCHI, 2013) como meios de amadurecimento da compreensão do objeto proposto.

Para Barbosa (2007), o Modelo de Jornalismo em Base de Dados (MJBD) classifica as bases de dados como uma ruptura no procedimento de criação das notícias jornalísticas e, portanto, um protocolo para o fazer jornalístico em rede na contemporaneidade influenciando nos formatos visuais das notícias. O uso de bases de dados para a produção jornalística não é recente, especialmente para o jornalismo investigativo. Entretanto, as próprias bases de dados tornam-se cada vez mais um meio e um produto jornalístico, ao invés de apenas um complemento a informação (Ibidem, 2007). Cada vez mais os principais jornais estão investindo em equipes específicas para esse tipo de trabalho e em produções de bases de dados interativas que possuam um caráter jornalístico narrativo. De acordo com o pesquisador e professor da Escola de Jornalismo de Medill, na Universidade de Northwestern, Rich Gordon, no artigo *Data as journalism, journalism as data*, de 2007, a hierarquia para os produtos do jornalismo de base de dados se organiza em cinco níveis, definidos conforme o grau de complexidade do uso que fazem das bases de dados. Esses níveis seriam:

1º Entrega de dados – as organizações de notícias obtêm os dados e os tornam disponíveis de uma forma navegável, mas não há reportagem ou interação para o usuário;

2º Pesquisa de dados – usuários encontram informações relevantes ao digitar entradas em uma caixa de texto;

3º Exploração de dados – é disponibilizada uma caixa de busca na própria página da base de dados permitindo uma fácil exploração dos dados;

4º Visualização de dados – linhas e colunas geralmente não são a forma mais eficaz de apresentar dados. Para muitos bancos de dados, a coisa mais valiosa que uma empresa de notícias pode fazer é fornecer uma maneira para as pessoas visualizarem o que os dados mostram. As abordagens mais óbvias envolvem mapeamento para bases de dados que tem um elemento geográfico, mas as possibilidades de visualização de

dados vão muito além de um mapeamento;

5º Experiência em dados e narrativas – união entre reportagem tradicional e narrativa com as capacidades do desenvolvimento de bases de dados e interação com o usuário.

Duas teses doutorais da Escola de Comunicações e Artes da Universidade de São Paulo contribuem para iluminar cientificamente o objeto visualização de dados: a primeira ao abordar o conceito de formato e a segunda ao abordar o sistema narrativo. Em 2012, na conclusão da tese *Formato, condição para a escrita do jornalismo digital de bases de dados : uma contribuição da semiótica da cultura*, a professora Daniela Osvald Ramos, apresentou um estudo teórico, à luz da semiótica da cultura, sobre como os formatos no jornalismo digital servem de estrutura para as linguagens digitais, entre estes formatos, a visualização de dados. Já a pesquisadora Daniela Bertocchi, em sua tese doutoral intitulada *Dos Dados aos Formatos – um modelo teórico para o design do sistema narrativo no jornalismo digital*, defendida no início de 2014, concordou com a noção de formato de Ramos e apontou que o jornalismo guiado por dados diz respeito ao processo jornalístico que vai da captura, passando pela curadoria e resultando na visualização dos dados em formatos. Defendeu que criar e manipular dados configura-se como o primeiro ato da antenarração. “Antenarrar dados insere-se numa “cultura de *software*” e dialoga diretamente com o jornalismo de dados” (BERTOCCHI, 2013, p. 114). É um tipo de formato que recorre à utilização de vários *softwares* de processamento de dados, *softwares* invisíveis na interface que será acessada pelo usuário, mas onde são usadas ferramentas de visualização como o *ManyEyes*²⁵, a *Your.FlowingData*²⁶, a *Visual.ly*²⁷, entre inúmeras outras. Em suma, são ações que tem por objetivo tornar humanamente mais compreensível e apreensível um conjunto complexo de dados.

As transformações vivenciadas fazem parte de um ecossistema²⁸ social que envolve a todos num processo de comunicação extremamente ágil e facilitado pelos dispositivos

²⁵ O *ManyEyes* foi desenvolvido pela IBM e é usado para apresentar de forma visual o conteúdo. Disponível em: <http://www-958.ibm.com/software/analytics/manyeyes/>.

²⁶ A *Your.FlowingData* é utilizada para coletar dados através do Twitter, disponível em <http://your.floatingdata.com>.

²⁷ A *Visual.ly* é usada para analisar uma extensa quantidade de dados a partir do padrão do retângulos alongados (Squarified). Disponível em: <http://visual.ly>.

²⁸ Partimos da noção de ecossistema do alemão Ernest Haeckel (1869) que introduziu o termo “ecologia”, de forma inédita ao abordar a relação entre ambientes e seres vivos. Aproximamos este pensamento dos estudos de mídia ao abordar a relação entre organismos, fatores ambientais e interações entre esses elementos (CANAVILHAS, 2010).

tecnológicos e móveis (principalmente se falarmos de acessos via celulares, geolocalizadores, etc). Mudam-se as redações, os processos sociais de comunicação e as formas de interação. Neste contexto também será feita uma reflexão sobre as empresas informativas e os jornalistas que devem estar atentos as tendências e as transformações na atualidade, sendo o primeiro item proposto pela pesquisadora e professora da Escola de Comunicação e Artes da Universidade de São Paulo Elizabeth Saad (2003) o de maior relevância para este projeto de pesquisa:

- 1o) Linguagens, formatos, conteúdos: diferenças narrativas da notícia on-line; configuração de um novo ‘veículo’;
- 2o) Estratégias, negócios e estruturas empresariais: o jornal passa a ser parte de uma estrutura integrada de negócios; convergência dos meios de produção.
- 3o) Emergência das mídias sociais digitalizadas: papel da blogsfera; jornalismo cidadão; redes sociais, interação por avaliação e recomendação, comunidades. (SAAD, 2003, p.142).

Integrado a lógica da sociedade atual, a proposta teórica de estudo deste projeto é também complexa, com características interdisciplinares e desafiadoras. Consideramos relevante na tentativa de observar um jornalismo, reflexo de uma sociedade, que se articula e se representa cada vez mais através de números e imagens.

Objetivos

Objetivo geral:

Apresentar o estado, até o presente momento, no campo da visualização de dados no jornalismo brasileiro. Jornalismo que engloba o cruzamento de banco de dados, as construções de mapas, de cartografias, de infográficos dinâmicos em narrativas diferenciadas. O que para nós fornece ao usuário o entendimento de temas complexos de relevância social, de forma visual, sintonizando o jornalismo com as novas necessidades de compreensão informativa da sociedade contemporânea;

Objetivo específico:

Observar como (fontes, uso de *softwares*, metodologias de análise de dados) o jornalismo de dados brasileiro vem utilizando a visualização de dados na produção

noticiosa. Como se insere na prática e na narrativa do jornalismo de dados proporcionando diferenciais relevantes (pautas de interesse público, que envolvem maior criticidade, aprofundamento, correlações e contextualização dos conteúdos no *Big Data*).

Objetivo teórico:

Contribuir para os estudos da Comunicação contextualizando as teorias com o cenário atual proporcionado pelos avanços tecnológicos na coleta e no auxílio ao tratamento de grandes volumes de dados no jornalismo.

Objetivo prático:

Mostrar a construção do conteúdo noticioso a partir da visualização em base de dados, questionando o papel do jornalista neste cenário onde dados podem ser somados, comparados e integrados entre si, complementando os grandes blocos de texto que o computador só consegue mostrar, mas não correlacionar, sem ser programado para isso.

Mapa da dissertação

A dissertação está dividida em cinco capítulos e considerações finais. No primeiro capítulo são apresentados os conceitos sobre visualização de dados e a distinção desta prática diante de outras formas visuais, pontuando a relação entre o jornalismo pós-industrial e o momento *Big Data*.

O segundo capítulo foca no cenário brasileiro do uso de dados. Um panorama do cenário jornalístico nacional, exemplificado com práticas que envolvem veículos de grande mídia, iniciativas colaborativas e estudos desenvolvidos em laboratórios acadêmicos, finalizando com as discussões em torno de regulamentações do tráfego de dados na *web*.

No terceiro capítulo visamos o aprofundamento nas visualizações de dados apontando

em que momento esta prática se encontra dentro da produção noticiosa. Para tanto, abordamos os conceitos de narrativa como sistema, de formato e designer no jornalismo. Abordamos as inovações da narrativa digital jornalística e quais os principais impactos para as empresas de mídia. Finalizando com as questões de credibilidade e valor jornalístico diante do cenário cultural contemporâneo.

No quarto capítulo são abordados os procedimentos metodológicos percorridos para tentar resolver os questionamentos deste estudo. A pesquisa multimetodológica é explicitada, assim como os critérios de observações adotados para o olhar ao blog Estadão Dados.

O quinto capítulo tem por objetivo exemplificar a prática da visualização de dados a partir dos conteúdos trabalhados no blog Estadão Dados. Parte-se de uma descrição geral do Blog e peculiaridades de cada sessão, para daí entrar nos processos incluídos nas peças jornalísticas.

Por fim temos as considerações finais que retomam as perguntas do nosso objeto: o que é a visualização de dados no jornalismo? Como ela se insere na prática e na narrativa do jornalismo de dados proporcionando diferenciais relevantes? Em suma, como o jornalismo vem utilizando a visualização de dados na produção noticiosa brasileira. Como acontece a captura, a curadoria dos dados, os processos de antenarrativa, e por fim, a visualização em um formato a ser acessado pelos usuários nas interfaces digitais. Assim como quais as influências presentes nesta prática. Por que trabalhar com a visualização de dados no jornalismo? Seria uma possibilidade de justificar a credibilidade? Como anda a expertise dos jornalistas para trabalhar nesta área?

Temos, portanto, esquematicamente pilares conceituais e estruturais.

- Pilares conceituais do estudo:
 - a. dados – dados como fonte: extraídos, lidos, correlacionados, *data mining*, *data scrap*;

- b. jornalismo – *data-driven journalism* ou jornalismo guiado por dados - reportagem tradicional + narrativas inovadoras com base de dado + convergências de diferentes áreas = forma diferenciada de jornalismo digital.
 - c. visualização de dados – diferenças entre visualização e visualização de dados, formas de visualização para a equipe jornalística e para os leitores;
 - d. narrativa – onde fica a visualização de dados no sistema narrativo (antenarrativa e narrativa/formato). Instâncias pré e pós (modus operandi) do usos de dados na narrativa.
 - e. expertise profissional – o que o profissional precisa saber para atuar.
- Pilares de modus operandi:
 - a. Estrutural básico: recursos (investimentos próprios ou recursos disponíveis gratuitamente); trabalho em equipe (interdisciplinar); tempo; local de publicação; leitores usuários;
 - b. Ferramental: dados (base de dados, fontes públicas, privada, fonte própria); tecnologias (criação e uso de ferramentas); análise e correlações (aproximações com outros recursos).

I. VISUALIZAÇÃO DE DADOS NO JORNALISMO BIG DATA

Resumo: Este capítulo tem por objetivo apresentar os conceitos sobre visualização de dados e a distinção desta prática diante de outras formas visuais, pontuando a relação entre o jornalismo (pós-industrial) e o momento *Big Data*. Tem como meta a contextualização desta prática diante dos estudos contemporâneos que remetem ao paradigma Jornalismo Digital de Base de Dados chegando ao jornalismo de dados. Temas que vão permitir aprofundar os estudos e a compreensão sobre a visualização de dados nos capítulos subsequentes.

Palavras-chave: Jornalismo pós-industrial, Big Data, Jornalismo Digital de Base de Dados, jornalismo de dados, visualização de dados

“Como podemos nadar en ese mar de información? No tenemos que beberlo todo, tenemos que aprender a nadar”. Pierre Levy (2015).²⁹

Dado e informação

A visualização de dados está cada vez mais presente nos meios de comunicação e iniciativas jornalísticas que buscam conteúdos com maior aprofundamento e contextualização. Ao visitar as páginas *web* de jornais nacionais e internacionais é possível perceber o crescimento das informações visuais com seções exclusivas de dados, assim como as já presentes sobre infografias. As iniciativas que reúnem imagens relacionadas a um grande volume de dados ocorrem nas práticas de mercado, laboratórios acadêmicos, assim como em mídias independentes. Diversas empresas jornalísticas tem se aperfeiçoado e disponibilizado na rede informações aprofundadas sobre situações contemporâneas. O jornalismo de dados³⁰ aparenta preencher as lacunas de conteúdos mais densos e contextualizados aos anseios de uma sociedade interconectada, que cada vez mais se representa a partir de imagens e números.

²⁹ Fonte: Twitter @Plevy, do dia 14 de maio de 2015, disponível em: <https://twitter.com/plevy>.

³⁰ Abordaremos o conceito sobre jornalismo de dados no decorrer deste capítulo.

Transcodificados para a linguagem computacional, os dados na *web* tornam-se elementos quantitativos e manipuláveis possibilitando diferentes formas de visualização a partir de ferramentas da computação que agilizam o trabalho de fazer correlações entre os dados.

Chegamos a difícil tarefa, neste capítulo, de apresentar definições e nomenclaturas encontradas ao longo deste estudo. Elas se fazem necessárias para contextualizar o cenário contemporâneo de uso de dados e para dar continuidade as discussões sobre o jornalismo e visualização de dados em capítulos posteriores. Conceitos sobre o que são dados, informações, design, arquitetura da informação e design da informação nos levam a ver melhor os conflitos nestas definições e nos posicionarmos enquanto pesquisadores. O primeiro ponto a ser trabalhado é exatamente a distinção entre informação e dado, para consequentemente distinguirmos entre visualização de dados e visualização da informação.

Informação enquanto conceito carrega uma diversidade de significados que vão do uso cotidiano ao técnico. De acordo com o Dicionário Contemporâneo da Língua Portuguesa, mais conhecido no Brasil por *Dicionário Caldas Aulete*, informação vem do latim *informatio, onis* (delinear, conceber ideia), significa dar forma ou moldar algo³¹. Já na *Wikipédia*³², informação é o resultante do processamento, manipulação e organização de dados que representam uma modificação que pode ser quantitativa ou qualitativa no conhecimento do sistema que o recebe. Seja este conhecimento humano ou maquínico³³. Para Claude Shannon (1949), pai da teoria matemática da informação, *informação* é aquilo que justifica, logicamente, uma alteração ou o reforço de uma representação ou estado de coisas. Acrescenta algo a uma representação e estas representações podem ser explícitas a exemplo de um mapa ou implícitas como em uma atividade orientada para o receptor. O matemático estadunidense Norbert Wiener (1970), conhecido como o fundador da cibernética³⁴, define informação como “o termo que designa o conteúdo daquilo que permutamos com o mundo exterior quando nos ajustamos a ele, e que faz com que nosso

³¹ Disponível em: <http://www.aulete.com.br/informa%C3%A7%C3%A3o>. Acesso em 2 de março de 2016.

³² Incluímos nesta pesquisa algumas referências da *Wikipédia*, como forma de contextualização, por acharmos relevante as definições que são visitadas e *editadas* por qualquer pessoa.

³³ Disponível em: <https://pt.wikipedia.org/wiki/Informa%C3%A7%C3%A3o>. Acesso em 02 de março de 2016.

³⁴ Cibernética é o estudo dos autocontroles encontrados em sistemas estáveis, sejam eles mecânicos, elétricos ou biológicos. Foi Wiener quem visualizou que a informação como uma quantidade era tão importante quanto a energia ou a matéria. Fonte: <http://pt.wikipedia.org/wiki/Informação>. Acesso em: 26 de fevereiro de 2015.

ajustamento seja nele percebido” (WIENER, 1970, p. 17). No Manual de Teoria da Comunicação (2007), o professor e pesquisador português Joaquim Paulo Serra distingue a palavra e o conceito antes e após Claude Shannon. De forma sucinta, a informação para Serra é o resultado do processamento, manipulação e organização de dados que representa uma modificação (quantitativa ou qualitativa) no conhecimento do sistema que a recebe.

Os dados, por sua vez, são registros e quando adquirem sentido interpretativo se transformam em informação. “Dessa forma, percebemos um fluxo crescente de interpretação entre dado, informação e conhecimento, nessa ordem” (RIBEIRO, 2009, p. 32). De acordo com o designer Nathan Shedroff (1999), em um contexto de mídias eletrônicas e digitais não é possível considerar como informação o vasto número de estímulos, de dados disponíveis hoje na *web*. Esses dados são apenas matéria bruta, utilizada em construções comunicativas. Os dados precisam ser organizados, transformados e apresentados para terem significado e valor interpretativo. Neste sentido é possível compreender o caráter de apropriação e manipulação de dados facilitados pela linguagem digital. De acordo com o dicionário de língua portuguesa Michaelis, dado é um conjunto de material ou informações que estão disponíveis para análise. É também uma “representação de fatos, conceitos e instruções, por meio de sinais de uma maneira formalizada, possível de ser transmitida ou processada pelo homem ou por máquinas”³⁵. Se buscarmos a definição na *Wikipédia* encontramos que dados são um conjunto de informações (quantitativas, qualitativas, categóricas ou indefinidas) podendo ser organizadas ou não³⁶.

Para a Escola de Dados³⁷, dado é um valor atribuído a alguma coisa. Sendo as duas principais categorias de dados os dados qualitativos (tudo que se refere à qualidade de algo. Uma descrição de cores, textura, uma descrição de experiências, uma entrevista. Tudo isso é dado qualitativo) e os dados quantitativos (dados que se referem a números, tamanho, preço, nota em uma prova, etc). No entanto, há outras categorias como a dos dados categóricos que são os que categorizam o item descrito); os dados

³⁵ Disponível em: http://michaelis.uol.com.br/moderno/portugues/definicao/dados%20_938422.html. Acesso em 15 de junho de 2016.

³⁶ Disponível em: <https://pt.wikipedia.org/wiki/Dados>. Acesso em 02 de março de 2016.

³⁷ Programa que no Brasil nasceu dentro do capítulo brasileiro da Open Knowledge e é parte de uma comunidade global que trabalha para capacitar organizações da sociedade civil, jornalistas e cidadãos para usar dados abertos. Disponível em: <http://escoladedados.org/>

discretos (que são os numéricos com brechas na sequência entre eles) e os dados contínuos (que são dados em que todos os valores são possíveis, não havendo brechas entre eles). O desafio repousa na interpretação e leitura dos dados, sejam pelas máquinas ou pelos humanos.

Computadores são bem diferentes de humanos. Pode ser extremamente difícil fazer computadores extraírem informações de determinadas fontes. Algumas tarefas fáceis para humanos ainda são difíceis de serem automatizadas com computadores. A interpretação de um texto apresentado como uma imagem é por exemplo um desafio para um computador. Se você quiser que o seu computador processe e analise os seus dados, ele tem que ser capaz de lê-los. Isso significa que os seus dados têm que estar estruturados, passíveis de serem lidos por computadores. Um dos formatos mais usados para a troca de dados é o CSV, sigla em inglês para valores separados por vírgula. (ESCOLA DE DADOS, 2013)³⁸.

Em resumo, dados pra nós são um conjunto de elementos quantitativas, qualitativas, categóricas ou indefinidas podendo ser organizadas ou não. Cada ciência vai tratá-los de acordo com características distintas e há diversas atividades que podem ser aplicadas aos dados e as informações, entre elas destacamos o design. A noção de design é fundamental neste trabalho, uma vez que o design é que torna a operação da escrita digital possível. Em concordância com Daniela Ramos (2011), criar um design significa planejar, desenvolver uma série de critérios e delinear um roteiro, de forma intencional, para chegar a algum ponto, há alguma coisa³⁹. A atividade de aplicar design à informação (ou seja, a conjuntos de dados que já passaram por crivos interpretativos) é chamada, muitas vezes, de Arquitetura da Informação ou Engenharia do Conhecimento. A pesquisadora Giulia Giannella, em defesa de mestrado em 2014, na Universidade de São Paulo foi além desta definição e apontou que a Arquitetura da Informação contempla, essencialmente, questões relativas à organização dos dados e dos seus caminhos de acesso (navegação, usabilidade, etc.). Acrescentou que por sua vez o Design de Informação compreende, a organização das informações, as relações entre conteúdo, a apresentação visual e a sua relação com o usuário, sendo portanto mais completo e tendo em seu guarda-chuva a “visualização da informação”. Para ela, a visualização da informação é uma modalidade de organização e apresentação visual da informação que faz parte do universo maior do Design da Informação (GIANNELLA, 2014). A visualização da informação designa o

³⁸ Disponível em: <http://escoladedados.org/manual/cursos/o-que-sao-dados/>. Acesso em 15 de junho de 2016.

³⁹ Retomaremos esta definição de design no capítulo 3, onde iremos situá-lo no sistema narrativo.

processo cujo objetivo é “facilitar a compreensão de grande quantidade de dados e suas relações de semelhança, ordem e proporcionalidade”. (ibidem, 2014, p. 34).

Com mais de 20 anos de experiência em computação humana e design na interação com usuários, o consultor Peter Boersma (2004) defende o uso do termo *User Experience Design* (UXD) ao abordar o conjunto de disciplinas que compõem uma área maior formada pela Arquitetura da Informação, Design da Informação, Design de Interação, *User Interface*, Engenharia de Usabilidade, Visualização de Informação, Estratégia de Conteúdo e Design Gráfico, entre outras. O termo *User Experience Design* foi cunhado pelo pesquisador em ciências cognitivas Don Norman, em 1995, quando este ocupava o cargo de vice-presidente de tecnologia da *Apple*.

Em entrevista publicada em 2007 (Adaptive Path, 2007), Norman afirmou ter criado o termo porque considerava interface humana e usabilidade expressões muito limitadas. “Eu queria cobrir todos os aspectos da experiência de uma pessoa com o sistema, incluindo desenho industrial, gráficos, a interface, a interação física (...)”. (BERTOCCHI et al, 2015, p. 67).

Os estudiosos americanos que trabalham com a interação homem/computador Stuart Card, Jock Mackinlay e Ben Shneiderman, autores do livro *Readings in Information Visualization: Using Vision to Think* (1999), desenvolveram uma série de técnicas de visualização da informação. Com o mesmo desafio de situar os estudos, eles defendem que (CARD, MACKINLAY, SHNEIDERMAN, 1999, p. 7):

- Cognição externa: uso do mundo externo para alcançar cognição;
- Design da Informação: Design de representações externas para ampliar a cognição;
- Gráfico de dados: uso de representações visuais de dados abstratos para ampliar a cognição;
- Visualização: uso de representações visuais interativas, auxiliado por um computador, para ampliar a cognição;
- Visualização científica: uso de representações visuais interativas de dados científicos – normalmente baseados em aspectos físicos – auxiliado por um computador, para amplificar a cognição;
- Visualização da informação: uso de representações visuais interativas de

dados abstratos, auxiliado por um computador, para ampliar a cognição.

Já para o professor e consultor em *business intelligence* Stephen Few (2009), a visualização de dados abrange tanto a visualização da informação como a visualização científica. Sendo a visualização de dados uma guarda-chuva para os estudos na área. Como apontado, não há consensos entre as definições e nem sobre uma única área de pertencimento da visualização de dados. Ao longo da pesquisa encontramos, inclusive, outras nomenclaturas e definições desta relação que envolve dados, informação e visualização:

- visualização de dados e datavis (TUFTE, 1990);
- visualização da informação e infovis (GIANNELLA, 2014);
- visualização analítica e infografia analítica (CAIRO, 2008);
- infografia em base de dados ou infobases (RODRIGUES, 2009)

Neste estudo concordamos com as distinções entre dado e informação e entre visualização de dados e visualização da informação. **Adotamos aqui a visualização de dados como um processo, que será trabalhado em um jornalismo específico, o jornalismo de dados, que tem suas características próprias, mas que também se encontra em área fronteiriça de estudo.** Assim como Stefen Few (2009) é um guarda-chuva mais amplo, onde a partir da visualização dos dados ocorrem os seus desdobramentos⁴⁰.

No gráfico a seguir podemos observar que há um caminho lógico para a formação da estrutura de visualização de dados. Em um primeiro momento temos a exploração dos dados, que leva a *sense-making* ou sensemaking (grafia que adotaremos nesta pesquisa) que é o processo pelo qual as pessoas dão sentido à experiência, uma "construção de sentido", que leva a uma comunicação. Posteriormente há o uso das visualizações, que resultam em representações gráficas (mapas, infográficos, ferramentas interativas). Em Ciência da Informação, o termo *sensemaking* é mais frequentemente escrito como "tomada de sentido" e tem sido usado para reunir

⁴⁰ Em seu estudo, a pesquisadora Giulia Giannella (2014), optou pela neologismo “dispositivo infovis” ao se referir a este processo.

insights extraídos da filosofia, sociologia e ciência cognitiva (psicologia especialmente sociais). A informação não é a extremidade no processo de construção da compreensão. Quando os dados adquirem sentido interpretativo se transformam em informação (RIBEIRO, 2009).

Figura I. 1 – Visualização de Dados, 2013.

Visualização de dados		
Atividades	Exploração <i>Sense-making</i>	⇒ Comunicação
Tecnologias	Visualização da informação Visualização científica	⇒ Representação gráfica
Objetivos imediatos	Compreensão	
Objetivos finais	Boas decisões ¹⁴	

Fonte: FEW, 2009, p. 12.

Para nós, a visualização é um meio para jornalistas e designers analisarem e investigarem relações entre dados, como também para transformar em informação visualmente estruturada os dados. O que, como produto final midiático, objetiva facilitar a compreensão de conteúdos complexos para os leitores. Defendemos que a visualização de dados no jornalismo está em uma área fronteira de estudos porque também envolve outros caminhos acadêmicos, inclusive os já amplamente trabalhados na área da Comunicação, como o jornalismo visual. No entanto, não se restringe somente a ele, uma vez que envolve outros elementos, como defenderemos a partir de agora.

A visualização de dados no jornalismo contemporâneo

Os conceitos referentes a visualização de dados, como já inicialmente apontado, não são convergentes e seu uso é observado em diversas áreas como Engenharia, Estatística, Ciências da Computação, Design da Informação, Comunicação. Os dados disponíveis hoje dentro e fora da *web* vem de diferentes áreas, sejam coletas estatísticas, pesquisas científicas, cálculos matemáticos, transações financeiras e a

maneira de organizá-los pode facilitar os processos de compreensão da informação. Além dos tipos convencionais de dados, que são os números, as datas e as pequenas cadeias de caracteres, é cada vez mais comum a necessidade de suportar dados complexos, como imagens, áudios, vídeos, entre outros. Nas Ciências da Computação, a visualização de dados se caracteriza como a possibilidade de encontrar diferentes formas de ver grandes volumes de dados, disponíveis em diferentes extensões, para que possam ser analisados, observados (SOUZA & BUENO, 2013). Nem todas as informações geradas e armazenadas por computador se encontram em bancos de dados, em linhas e tabelas estruturadas. A maior parte não está estruturada, são informações guardadas em diversos documentos, e-mails, vídeos, conteúdos em redes sociais digitais, com extensões distintas, estruturados nas mais diferentes formas e com suas complexidades (TAKAI et al, 2005).

Os dados precisam ser organizados, transformados e apresentados para ter valor e significado. Esta característica, no ambiente digital, os torna ainda mais manipuláveis. Esta manipulação está cada vez mais atrelada a evolução da capacidade de processamento computacional. A chamada *data mining* ou mineração de dados, por exemplo, é uma área de investigação da Ciência da Computação que procura identificar padrões em um grande volume de dados digitais. Para isso recorre a algoritmos de aprendizagem baseados em técnicas estatísticas e de redes neurais. Esses algoritmos visam o processamento em bancos de dados em busca de evidências relevantes sobre tendências, comportamentos, relações invisíveis ou dificilmente identificáveis numa análise humana. Para o professor Aaron Quigley, presidente da Human Computer Interaction na Escola de Ciência da Computação da universidade escocesa, Saint Andrews, o processo de conversão dos dados, de uma origem abstrata e sem forma geométrica, para formas gráficas é conhecido como *visualização*. Quigley defende a habilidade natural humana de compreender a informação de maneira visual. Sendo assim, em um sentido amplo, a visualização é o processo de tornar visível o invisível (QUIGLEY, 2006).

No campo da Comunicação, a visualização de dados é entendida como “o aproveitamento do agenciamento entre as camadas de dados e metadados no âmbito do sistema narrativo, o qual privilegia a apresentação visual relevante e de impacto” (BERTOCCHI, 2013, p. 180). É um formato que recorre à utilização

de *softwares* de processamento de dados que não necessariamente aparecem ao usuário/leitor, uma vez que os resultados finais podem ser construções narrativas a partir dos dados obtidos, como infografias, mapas, entre outros⁴¹. A visualização de dados resulta em informações que são mais facilmente compreensíveis. Traduz a complexidade dos dados e suas correlações através de modelos visuais. As variações e os padrões de repetição são observáveis mais facilmente quando são traduzidos em formas geométricas. A pesquisadora Donna Cox (2006) defende que a construção de modelos de visualização de dados é um processo de mapeamento. Para Manovich (2004), ao representar diferentes dados usando o mesmo código numérico os computadores facilitam o mapeamento de uma representação em outra. Para o pesquisador Daniel Melo Ribeiro, em dissertação defendida na Pontifícia Universidade Católica de São Paulo (2009), o mapeamento estabelece uma relação direta entre a visualização de dados e a cognição, podendo criar vínculos através de metáforas visuais.

Tanto as metáforas linguísticas quanto as visuais são definidas com mapeamentos de um domínio de informação em outro. Cada um dos domínios constitui um sistema de crenças, onde os indivíduos veem, usam e interpretam as imagens de acordo com suas experiências, hábitos, preferências e bagagem cultural. A cultura moderna, como sabemos, está impregnada de imagens, gráficos e efeitos visuais. E as metáforas visuais impactam nossa compreensão da realidade cotidiana tanto quanto as metáforas linguísticas, em termos psicológicos e sociais. (RIBEIRO, 2009, p. 73).

A visualização compreende algumas características tais como o estabelecimento de relações topológicas entre os dados; representações que se atualizam a partir de lógicas sistêmicas; instrumento de descoberta, pois cada dado não é um elemento isolado, mas um nó conectado a vários pontos de diferentes redes. O que é preciso ter em mente é que ao utilizar modelos de visualização de dados são possíveis construções de sentido na medida que os dados são organizados de forma lógica, mapeados de uma maneira apropriada, contextualizados e transformados em informação. E no que difere, por exemplo, do jornalismo visual?

De acordo com o professor da Universidade de São Paulo Luciano Guimarães (2012),

⁴¹ O jornalista americano Josh Stern publicou um post no seu blog intitulado *Groundswell* com as reportagens que foram destaque no jornalismo em 2014. Elas utilizaram *storytelling*, recursos de áudio, vídeo, satélites, sensores e imersão. Exemplos interessantes que envolvem infográficos e mapas em narrativas originais. O post *The Best Online Storytelling and Journalism of 2014*, está disponível em <http://stearns.wordpress.com/2015/01/05/the-best-online-storytelling-and-journalism-of-2014/>. Acesso em 12 de janeiro de 2015.

o jornalismo visual abrange tanto as instâncias da Comunicação Visual como do Jornalismo, onde são criadas formas de representação do conteúdo informativo, resultando, na maioria das vezes, em manifestações texto-visuais. A prática existe desde o aparecimento do jornal e evolui com os recursos tecnológicos disponíveis, os resultados são desde charges, ilustrações, fotografias até os infográficos, entre outras práticas. No caso da visualização de dados, o que a diferencia de outras práticas é a sua origem e apuração que requer diferentes expertises no fazer jornalístico que emergem a partir de congruências e manuseio de grande volume de dados. Nos concentramos, prioritariamente, no fator tecnológico como um dos pré-requisitos para estudar as peças jornalísticas que serão exploradas no capítulo 5. Isso se dá pela demanda do uso de diferentes tecnologias e análises estatísticas diante da grande quantidade de dados, neste cenário contemporâneo do Big Data⁴².

Em concordância com o Manual de Jornalismo de Dados⁴³ (GREY et al. 2012), ao contrário de outros recursos visuais, como a fotografia e o vídeo, a visualização de dados está enraizada em fatos mensuráveis. Pode ser eficaz tanto para apresentar notícias, transmitindo rapidamente informações pontuais como a localização de um acidente e o número de vítimas, como para reportagens, nas quais pode aprofundar um tema e oferecer uma nova perspectiva sobre algo familiar. A visualização de dados e as associações estéticas que fazem parte dela, podem até se tornar uma referência cultural, assim como a representação de divisões políticas em um processo eleitoral. Para o professor de visualização de dados e infografia da Escola de Comunicação da Universidade de Miami, Alberto Cairo (2012), a informação desestruturada é sinônimo de realidade, do mundo que nos rodeia e de suas complexidades. Os fenômenos passíveis de serem observados se constituem como informação. Os dados por sua vez são registros de observações, se codificam como símbolos, sejam números ou palavras, e descrevem e representam a realidade. A informação estruturada ou semântica diz respeito aos dados que podem ser entendidos, sendo portanto, um segundo nível de codificação, que consiste em representar os registros de forma que se tornem compreensíveis. Para ele, o consumo de informações contextualizadas pode levar cada vez mais ao aumento do

⁴² Muitos estudos envolvendo este grande volume e variedade de dados vem sendo chamado de Ciência de Dados. Para o consultor Nuno Moura Pires (2014), a Ciência de Dados procura descobrir conhecimento, a partir de uma quantidade grande e pesada de dados, para a tomada de decisões e para fazer previsões. Não é simplesmente a interpretação de números. Concordamos com essa visão, mas para não adentrar em outra discussão iremos nos referir nesta pesquisa apenas por Big Data.

⁴³ Disponível em: <http://datajournalismhandbook.org/pt/index.html>. Acesso em 30 de janeiro de 2015.

conhecimento, onde o leitor assimila o que lhe é apresentado, comparando e mesclando com sua própria memória e experiência (CAIRO, 2012).

Visualização é aquela tecnologia plural (isto é, disciplina) que consiste em transformar dados em informação semântica – ou em criar as ferramentas para que qualquer pessoa complete por si só este processo – por meio de uma sintaxe de fronteiras precisas e constantes evoluções baseadas na natureza arbitrária e abstrata (não figurativas: textos, estatísticas, etc). (CAIRO, 2012, p. 38).⁴⁴

Neste sentido, torna-se interessante pontuarmos **a distinção, ao nosso ver, entre a visualização de dados e a infografia. Comumente elas caminham juntas nas peças jornalísticas, mas fundamentalmente e operacionalmente há distinções.** A infografia consiste em apresentar informações, em reduzir, editar e eleger os dados relevantes, eliminando dados secundários. Nelas são utilizadas formas gráficas para que os dados sejam compreensíveis pelo público leitor. Já a visualização de dados é a disciplina que se ocupa de criar ferramentas visuais, muitas vezes interativas, para que uma determinada audiência, geralmente composta por pesquisadores e acadêmicos, possam estudar e explorar os dados. No jornalismo, pode ser uma etapa para se trabalhar os dados e posteriormente apresentá-los, em diferentes designs, para o leitor, inclusive em um infográfico.

Explicado de outra forma: em uma visualização o usuário ocupa o lugar de editor: elege quais dados quer ver e como quer vê-los para analisá-los com maior comodidade. Apresentação e exploração são as funções que, de acordo com esta corrente teórica, caracterizam e separam a infografia da visualização. (CAIRO, 2011)⁴⁵.

Para nós, como já mencionado, a visualização de dados é um termo guarda-chuva que envolve um processo que determina a forma de apresentação do produto final em um trabalho jornalístico⁴⁶.

⁴⁴ “Visualización es aquella tecnología plural (esto es, disciplina) que consiste en transformar datos en información semántica — o en crear las herramientas para que cualquier persona complete por sí sola dicho proceso— por medio de una sintaxis de fronteras imprecisas y en constante evolución basada en la conjunción de signos de naturaleza icónica (figurativos) con otros de naturaleza arbitrária y abstracta (no figurativos: textos, estadísticas, etc.)”. (CAIRO, 2012, p. 38).

⁴⁵ “Explicado de otra forma: en una visualización el usuario ocupa el lugar del editor: elige *qué* datos quiere ver y *cómo* quiere verlos, para analizarlos con mayor comodidad. Presentación y exploración son las funciones que, según dicha corriente teórica, caracterizan y separan a la infografía de la visualización” (CAIRO, Alberto, 2011, tradução nossa).

⁴⁶ Como contextualização, há distinções entre a visualização de informação e infográficos⁴⁶. Para o pesquisador Robert Kosara (2010), os infográficos são, em teoria, produzidos manualmente e utilizados para conjuntos específicos de dados, enquanto visualizações de informação são desenvolvidas de modo mais automático por ferramentas que possibilitam aplicar uma mesma forma gráfica a diversos conjuntos de dados. Giannella (2014) discorda de Kosara sobre a tecnologia de produção ser

Exemplificando

Partindo para exemplos práticos, toda visualização de dados pode ser representada como converter-se em uma ferramenta passível de manipulação pelo leitor. No entanto, “a chave dos melhores gráficos consiste em ter um equilíbrio entre apresentar e facilitar certos aprofundamentos”⁴⁷ (CAIRO, 2011). Os melhores gráficos conseguem extrair significados e tendências vinculados ao cotidiano das pessoas. No Brasil, as produções em visualização de dados e peças infográficas, entre outros formatos em base de dados, são comparativamente mais recentes e em menor escala. Houveram iniciativas, por exemplo, como o Infolide⁴⁸, evento anual realizado em São Paulo e que evidenciava produções infográficas nacionais. Em 2013, seis trabalhos brasileiros⁴⁹ foram premiados no Malofiej, que se destaca como uma das maiores premiações mundiais de infografia⁵⁰. O jornal O Estado de S. Paulo chegou a conquistar medalha de ouro na categoria Portfólio, com a série sobre os Estádios da Copa de 2014⁵¹. Já conhecido o projeto InfoAmazônia⁵², do jornalista Gustavo Faleiros, esteve entre os finalistas. Em 2015 o blog Estadão Dados, do Jornal O Estado de S. Paulo foi o grande vencedor da 60ª edição do ExxonMobil 2015 (antigo Prêmio Esso de Jornalismo)⁵³, com a matéria Farra do Fies, onde foram correlacionados dados de mais de 42 mil alunos com bases de dados do Censo de Educação Superior e do Portal de Transparência, do Governo Federal (este caso será melhor explorado no 5º capítulo deste estudo).

O jornal *The New York Times* vem se destacando em coberturas de dados com o uso de ferramentas próprias e em narrativas diferenciadas e interativas. A matéria

determinante para a distinção entre infografia e visualização da informação. Para ela, a infografia é uma prática ampla que gera produtos jornalísticos texto-visuais, customizados ou com base em *templates*, criados para fins predominantemente informativos usando conhecimentos teóricos e práticos do campo da Comunicação e do Design. Já a visualização da informação, seja ela interativa ou não, constitui uma modalidade de representação gráfica que visa elucidar um fato, um ponto de vista objetivo, consistindo em processo de ampliação da cognição e do conhecimento. Ainda de acordo com Giannella, há infográficos interativos que se apropriam das técnicas da visualização da informação para organizar, cruzar e apresentar relações entre dados constituindo-se como um dispositivo midiático de características específicas que tanto demonstra mediação jornalística do conteúdo como permite ao usuário certa autonomia na construção do percurso narrativo por meio de recursos interativos. Neste meio existiria também as infobases, que tem a base de dados e uso de aplicativos de *softwares* baseados na *web* (APIs⁴⁶) em sua constituição, que utilizam dados armazenados em outros diretórios.

⁴⁷ “la clave de los mejores gráficos consiste en lograr un equilibrio entre presentar y facilitar cierta profundización” (CAIRO, 2011, tradução nossa).

⁴⁸ Disponível em: <http://www.infolide.com>. A última edição do evento aconteceu em 2013.

⁴⁹ Ver: <http://visualoop.com.br/5250/os-infograficos-brasileiros-premiados-no-malofiej-21>

⁵⁰ <http://www.malofiejgraphics.com>

⁵¹ Disponível em: <http://www.estadao.com.br/infograficos/maracana-fica-mais-moderno-sem-abrir-mao-de-sua-historia-esportes,208362>

⁵² Disponível em: <http://infoamazonia.org/>

⁵³ <http://politica.estadao.com.br/noticias/geral,pelo-segundo-ano-consecutivo--estado-vence-o-principal-premio-de-jornalismo-no-brasil,1782259> ou <http://politica.estadao.com.br/noticias/geral,jornalistas-do-estado-recebem-no-rio-premio-esso-de-jornalismo--imp-,1795558>

*Reshaping New York*⁵⁴, é um exemplo. Ela foi premiada na categoria melhor visualização de dados no *Data Journalism Awards 2014*⁵⁵, em Barcelona.

Figura I. 2 – Nytimes.com, matéria Reshaping New York, 2013.

Fonte: The New York Times, 2013.

A matéria trouxe inúmeras informações sobre o boom de novas habitações em *Nova York*, mesmo no período de recessão. Os dados mostraram a mudança nos bairros nos últimos anos, incluindo a presença de edifícios e ciclovias. É uma leitura imersiva, interativa, repleta de fotos e que retrata as modificações de doze anos em que o prefeito Bloomberg esteve no comando. A partir da matéria é possível ver o crescimento maciço na construção residencial, o rezoneamento de todos os cinco distritos para facilitar o crescimento econômico, a adição de 450 milhas de ciclovias,

⁵⁴ Disponível em: <http://www.nytimes.com/newsgraphics/2013/08/18/reshaping-new-york/>. Acesso em 26 de fevereiro de 2015.

⁵⁵ O evento é um dos principais agregadores de publicações inovadoras. É organizado pela Global Editors Network, uma comunidade multi-plataforma de editores-chefes e inovadores de mídia comprometida com o jornalismo sustentável e de qualidade. Disponível em: <https://www.globaleditorsnetwork.org/programmes/data-journalism-awards/>

de dezenas de praças de pedestres e muitas outras informações referentes a cidade. Há fotos de “antes/depois”, edifícios e bairros clicáveis e recursos interativos que permitem que você faça diversos caminhos na busca de informações.

Já o vencedor na categoria *Knight Award for Public Service*, do *Online Journalism Awards* (OJA)⁵⁶, em 2014, foi a publicação *Inocents Lost*, do jornal *Miami Herald*⁵⁷. O material, fruto de um intenso trabalho de investigação, reúne informações sobre o aumento das mortes de crianças após o Estado americano da Flórida reduzir as proteções para as crianças em lares problemáticos. Entre as fontes, relatórios policiais, autópsias, entrevistas e busca nos últimos cinco anos dos homicídios registrados pelo *Department of Children & Families*. Foi criada uma base de dados de cada morte, com as fichas policiais, mapas e relatos que pôde ser acessada pelo leitor.

Figura I. 3 - Miamiherald.com, matéria *Inocents Lost*, vencedora do OJA 2014.

Fonte: Miami Herald, 2014.

⁵⁶ A premiação que existe desde 2000 foi criada pela associação de jornalistas *Online News Association* (ONA)⁵⁶ com o objetivo de inspirar e estimular os jornalistas a desenvolverem habilidades e competências que lhes permitam fazer uso das ferramentas digitais para melhor contar as suas histórias. Disponível em: <http://journalists.org/awards/2014-awards/>

⁵⁷ Disponível em: <http://media.miamiherald.com/static/media/projects/2014/innocents-lost/>

O *Malofiej Awards*, como já mencionado, é uma das maiores premiações mundiais de infografia que acontece desde 1993. O infográfico *State Gun Laws Enacted in Year Since Newton*⁵⁸, também do *The New York Times*, foi o vencedor, em 2014, em duas categorias: reportagem impressa e prêmio *Peter Sullivan /Best of Show*. A reportagem aborda o porte de armas após o tiroteio na *Sandy Hook Elementary School*, em *Newtown, Connecticut* e que após este acontecimento em quase todos os Estados pelo menos uma nova lei de armas entrou em vigor. Os dados foram compilados pelo Centro de Direito para Prevenir a Violência Armada (em inglês, *Law Center to Prevent Gun Violence*). Há informações de cada Estado, com linhas do tempo. Nas correlações, a maioria das aprovações foram em Estados controlados pelos republicanos, que tendem a aprovar leis que diminuem a restrição quanto ao porte de arma. Ocorrendo, comparativamente, o inverso nos Estados democráticos.

Figura I. 4 – Nytimes.com, matéria State Gun Laws Enacted in Year Since Newton, 2013.

State Gun Laws Enacted in the Year Since Newtown

About **1,500** state gun bills have been introduced since the Newtown massacre.

178 passed at least one chamber of a state legislature. **109** have become law.

In the 12 months since the mass shooting at Sandy Hook Elementary School in Newtown, Conn., almost every state has enacted at least one new gun law, according to a database compiled by the Law Center to Prevent Gun Violence. Nearly two-thirds of the new laws ease restrictions and expand the rights of gun owners. Most of those bills were approved in states controlled by Republicans. Those who support stricter regulations won some victories — mostly in states where the legislature and governorship are controlled by Democrats — to increase restrictions on gun use and ownership. Select categories from the table below to see all gun bills that passed at least one chamber of a state legislature.

Fonte: The New York Times, 2013.

⁵⁸ <http://www.nytimes.com/interactive/2013/12/10/us/state-gun-laws-enacted-in-the-year-since-newtown.html>

A partir da breve descrição dos exemplos é possível perceber que ocorre a formação de distintas base de dados e a correlação entre elas. As formas de levar estas informações aos leitores também é diversa, mas todas conseguem aprofundamentos e reflexões importantes, permitindo que o usuário/leitor se aproprie e navegue por diferentes caminhos. Os conteúdos também tornam-se base para outras matérias e passam a ser referência na busca de informações. Os projetos que se tornam permanentes recebem atualizações constantes. Alguns em tempo real, linkados as próprias bases de dados de polícias e outras instituições envolvidas.

O que queremos discutir de forma aprofundada neste estudo é a forma de execução, os processos de decisão, elementos que ocorrem em um *backend* jornalístico que nos capítulos seguintes serão melhor explorados. A leitura dos conteúdos não nos permite entender o todo, os recursos e as dificuldades envolvidas no fazer, mas já nos dimensionam os caminhos do jornalismo de dados.

O jornalismo na era Big Data

Sociedade, tecnologia, comunicação, economia, há pontos-chave que devem ser trabalhados para se entender o contexto do jornalismo digital na contemporaneidade. Aspectos que refletem em mudanças na prática, nas necessidades de recursos, na forma de ver o produto jornalístico. Diferentes denominações tem sido utilizadas para nomear este momento sociocomunicacional (SAAD, 2014) como espaço público social (MIÈGE, 2009), esfera pública conectada (BENKLER, 2006), sociedade informacional (CASTELLS, 2003), sociedade da comunicação generalizada (VATTIMO, 1992), entre outras, mas seja qual for a denominação e suas variáveis, vivencia-se um ciberespaço híbrido, de linha tênue entre as esferas pública e privada. Estamos em um período de novas formas de conhecimento e de acesso a informações possibilitado pelas inúmeras experimentações no ambiente digital. Saber o que acontece no Brasil, no mundo e no círculo de amigos através das redes sociais é cada vez mais comum, por exemplo.

Vivemos uma era em que se convencionou chamar de *Big Data* onde há uma quantidade enorme de informações que os humanos produzem e não conseguem processar sem a ajuda da computação. Situação mediada por tecnologias e que demanda um contínuo desenvolvimento tecnológico por parte de governos, instituições e empresas privadas. ***Big Data*, portanto, é um termo que abrange o crescimento, a disponibilidade e o uso exponencial de informações e leva a questionamentos sobre como pode servir como base para a inovação, a diferenciação e o conhecimento sobre os mais diferentes aspectos da sociedade seja cultura, política, religião, saúde, habitação, entre inúmeras outras finalidades.**

O volume exponencial de dados advém do armazenamento de conteúdos ao longo dos anos, dados de texto em *streaming*⁵⁹ nas mídias sociais, dados criados por usuários, e-mails, vídeos, áudios, mercado de transações financeiras. Em 2011, o relatório "Big Data Analytics: Future Architectures, Skills and Roadmaps for the CIO", do International Data Corporation – IDC⁶⁰ (empresa de inteligência de mercado e consultoria nas indústrias de tecnologia da informação, telecomunicações e mercados de consumo em massa de tecnologia) já apontava que 80% dos dados de uma organização não eram numéricos, estavam disponíveis em outras extensões.

Há também uma tendência mundial de governos em aumentar a transparência dos dados disponibilizando cada vez mais informações referentes à administração pública. As iniciativas *open data* também tendem a ser mais frequentes, assim como a produção de conteúdos jornalísticos e seções em sites especificamente para disponibilizar dados produzidos, organizados e estruturados por profissionais através de APIs. No Brasil são diversas as iniciativas como o *Dados.gov.br*⁶¹, a *Lei de Acesso à Informação*⁶²; são inúmeras as formas de armazenamento, processamento, análises e visualizações de dados. Até recentemente as organizações se limitavam ao uso de subconjuntos de amostras dos seus dados e a análises simplistas porque os volumes absolutos de dados sobrecarregavam as plataformas de processamento. Cada vez há

⁵⁹ De acordo com o site de tecnologia da Globo.com, TechTudo, streaming é uma forma de transmissão instantânea de dados de áudio e vídeo através de redes. Por essa tecnologia é possível assistir a filmes ou escutar música sem a necessidade de fazer download, o que torna mais rápido o acesso aos conteúdos online (COUTINHO, Mariana (2013). Saiba mais sobre streaming, a tecnologia que se popularizou na web 2.0. Disponível em: <http://www.techtudo.com.br/artigos/noticia/2013/05/conheca-o-streaming-tecnologia-que-se-popularizou-na-web.html>. Acesso em: 25 de junho de 2016).

⁶⁰ Disponível em: <http://br.idclatin.com/default.aspx>

⁶¹ Disponível em: <http://dados.gov.br>

⁶² Disponível em: <http://www.acessoainformacao.gov.br/assuntos/conheca-seu-direito/a-lei-de-acesso-a-informacao>

mais necessidade de incorporar grandes volumes de dados na análise usando diferentes tecnologias como *Grid Computing*⁶³, processamento *In-Database*⁶⁴, análise *In-Memory*⁶⁵ ou *cloud computing*⁶⁶. Outro ponto é já aplicar a análise na entrada dos dados para determinar sua relevância com base no contexto. Estas análises podem ser usadas para determinar quais dados devem ser incluídos nos processos de análise e quais podem ser alocados em armazenamento de baixo custo de acesso posterior. De acordo com a empresa SAS (o nome vem de *statistical analysis system* ou sistema de análise estatística)⁶⁷, que oferece serviços na área de *business analytics software* e serviços, as organizações terão que ser capazes de aproveitar os dados relevantes e usá-los para tomar as melhores decisões, a exemplo do *Big Data Analytics*. Através destes estudos é possível (SAS, 2011):

- Determinar as causas de falhas, problemas e defeitos, investigando as sessões de usuário, logs de rede e sensores da máquina.
- Minerar dados de clientes para insights que impulsionam novas estratégias de otimização de campanhas, aquisição e retenção de clientes.
- Analisar milhões de SKUs (uma forma de código de barras, de localizador) para determinar os preços ótimos que maximizam os lucros e mínimo inventário.
- No Varejo, gerar cupons no ponto de venda com base em compras atuais e passadas, garantindo uma maior taxa de conversão.
- Enviar recomendações personalizadas para dispositivos móveis no momento certo, enquanto os clientes estão no local para aproveitar as ofertas.
- Recalcular carteiras de risco inteiras em questão de minutos e compreender as possibilidades futuras para mitigar o risco.
- Rapidamente identificar os clientes mais importantes.
- Analisar dados de mídias sociais para detectar novas tendências de mercado e mudanças na demanda.
- Usar a análise de fluxo de cliques e de mineração de dados para detectar comportamentos fraudulentos (SAS, 2011, tradução nossa)⁶⁸.

⁶³ O *grid computing* é um modelo computacional capaz de alcançar uma alta taxa de processamento dividindo as tarefas entre diversas máquinas, podendo ser em rede local ou rede de longa distância, que formam uma máquina virtual. Esses processos podem ser executados no momento em que as máquinas não estão sendo utilizadas pelo usuário, assim evitando o desperdício de processamento da máquina utilizada. Disponível em: http://en.wikipedia.org/wiki/Grid_computing. Acesso em 29 de janeiro de 2015.

⁶⁴ O processamento *In-Database* muitas vezes se refere a análise de banco de dados, a integração de sistemas para o armazenamento de dados. É um método muito utilizado por redes bancárias para a detecção de fraudes com cartões de crédito e gestão de risco em investimentos. Disponível em: http://en.wikipedia.org/wiki/In-database_processing. Acesso em 28 de janeiro de 2015.

⁶⁵ A análise *in-memory* é um recurso de armazenamento de grande volume de dados diretamente na memória principal dos servidores. A grande diferença deste modelo de processamento de dados para o convencional é que, anteriormente, estas informações eram arquivadas em discos rígidos, o que reduzia a agilidade na consulta pelas informações em alta escala, já que forçava a memória do computador a realizar estas consultas. Disponível em: <http://www.artsoftistemas.com.br/blog/entenda-como-o-in-memory-computing-e-essencial-para-aproveitar-as-oportunidades-do-mercado>. Acesso em 15 de janeiro de 2015.

⁶⁶ O *cloud computing* refere-se à utilização da memória e das capacidades de armazenamento e cálculo de computadores e servidores compartilhados e interligados por meio da Internet, seguindo o princípio do *grid computing*. Disponível em: http://pt.wikipedia.org/wiki/Computação_em_nuvem. Acesso em 28 de janeiro de 2015.

⁶⁷ http://www.sas.com/en_us/insights/big-data/what-is-big-data.html

⁶⁸ “Determine root causes of failures, issues and defects in near-real time, potentially saving billions of dollars annually. - Optimize routes for many thousands of package delivery vehicles while they are on the road. - Analyze millions of SKUs to determine prices that maximize profit and clear inventory. - Generate retail coupons at the point of sale based on the customer's current and past purchases. - Send tailored recommendations to mobile devices while customers are in the right area to take advantage of offers. - Recalculate entire risk portfolios in minutes. - Quickly identify customers who matter the most. - Use

Os desafios contemporâneos envolvem grande volume de dados, como armazená-los, como acessá-los, quais conteúdos são relevantes e como usá-los. Envolve as questões de público e privado, em um cenário em constante mudanças, aprendizado e aperfeiçoamento. Propõe-se o estudo da visualização de dados no jornalismo digital no momento de um jornalismo pós-industrial, não mais organizado consoante uma lógica industrial em cascata produtiva; mas com marcas mais complexas: mais atores atuantes, mais circularidade, mais algoritmos, mais inteligência artificial, mais computação em seu interior (BERTOCCHI, 2013, p. 10). As práticas jornalísticas se atrelaram, portanto, as constantes mudanças sociais e tecnológicas, tendo diversas fases e formas de apresentação de conteúdos desde o surgimento do online. Os modelos de negócios estão em adaptação, principalmente para serem sustentáveis num momento em que indivíduos, instituições, empresas ou qualquer organização possuem o poder de mídia.

De acordo com o pesquisador brasileiro Caio Tulio Costa (2014), em pesquisa feita juntamente com a *Columbia University Graduate School of Journalism*⁶⁹, as empresas jornalísticas estão pagando caro por terem descuidado dos investimentos em tecnologias, pela falta de foco no desenvolvimento de produtos jornalísticos no ambiente digital. “(...) a *web* mudou a forma como organizamos a informação – que veio daquele formato claro dos livros e dos jornais, delimitado e sólido, para algo de consistência parecida com o líquido, cujo fluxo é livre e de possibilidades ilimitadas (...)”. (COSTA, 2014, p. 84). A informação produzida por um jornal de qualidade pode ser recortada, retrabalhada, reescrita e compartilhada de forma absolutamente diferente, e até deformada, em relação ao seu conteúdo original. A problemática também se encontra em gerenciar as diferentes visões, principalmente as enraizadas em práticas consolidadas ao ambiente analógico e com profissionais que tem pouca experiência em processos digitais.

A indústria do jornalismo, seja por questões geracionais, seja por questões de negócio, seja por mera incompetência, levou muito mais tempo para entender o momento disruptivo pelo qual passa do que levaram, por exemplo, a indústria das telecomunicações, que soube recriar-se no ambiente sem fio; a indústria da música, que está se refazendo no ambiente digital; ou mesmo o varejo tradicional, que está se recuperando no comércio online. (COSTA, 2014, p. 61).

clickstream analysis and data mining to detect fraudulent behavior”. SAS (2011). Big Data: what it is and why it matters. Disponível em: <http://www.sas.com/en_us/insights/big-data/what-is-big-data.html> Acesso em: 28 jan. 2015.

⁶⁹ <https://journalism.columbia.edu/>

O negócio do jornalismo passa a ser mais do que conteúdo, mas também de distribuição. “Seu negócio é a prestação de serviços, com a integração vertical de conteúdo, reprodução e distribuição” (COSTA, 2014, p. 55). A questão não é somente como aumentar a audiência, “mas como fazer com que os internautas em rede se dediquem aos seus conteúdos nas outras plataformas. E, o mais importante de tudo, conseguir rentabilizar isso”. (COSTA, 2014, p. 85). As receitas dos veículos tradicionais ainda galgam caminhos para se consolidarem nas múltiplas formas de atuação no online⁷⁰. No entanto, para diversos autores do jornalismo pós-industrial, a dicotomia entre lucratividade e independência do jornalismo de qualidade não existe, uma vez que o jornalismo sempre foi subsidiado de alguma forma. Uma das características vivenciadas pelo jornalismo na atualidade, no chamado jornalismo pós-industrial, são os efeitos disruptivos da Internet e das tecnologias de combinar modelos de meios de comunicação num único canal. Neste contexto, o jornalista continua a fazer parte do sistema de informação; não mais como o seu principal ator, porém, não menos importante.

(...) a saída que se antevê para os modelos de negócios possíveis nesta nova realidade tem de levar em conta que o jornalismo deve continuar a expor a corrupção, chamar a atenção para as injustiças, cobrar políticos e empresas pelas promessas e obrigações assumidas, informar cidadãos e consumidores, ajudar a organizar a opinião pública, elucidar temas complexos e esclarecer divergências fundamentais – ou seja, os jornalistas não são apenas narradores de fatos (COSTA, 2014, p. 64).

Prestação de serviço, pautas aprofundadas, contextualizadas, apresentadas de acordo com as diferentes formas de cognição no ambiente digital. São diversos os desafios para a prática jornalística e para as empresas de mídia na contemporaneidade. Neste contexto, os profissionais não são necessariamente multitarefa, mas precisam ter clareza de diversas competências e saber trabalhar em equipe. Os trabalhos são cada vez mais criativos, complexos e unem inúmeros recursos como áudio, vídeo e texto em narrativas originais e não-lineares. Conhecer o leitor torna-se essencial em um trabalho que se torna cada vez mais de nicho. Isso requer um conjunto de outras

⁷⁰ Nós não vamos nos aprofundar neste estudo em modelos de negócio no jornalismo digital, mas há caminhos que estão sendo adotados pelos veículos tradicionais de mídia, entre eles os de conteúdo *paywall*. Depois de 13 anos registrando resultados negativos, os jornais *The Times* e *Sunday Times*, da *Times Newspapers*, por exemplo, apresentaram lucro no ano fiscal, encerrado em junho de 2014. É o primeiro resultado positivo desde a adoção do modelo de *paywall*, em 2010. Disponível em: <http://www.meioemensagem.com.br/home/midia/noticias/2014/12/03/Times-e-Sunday-tem-lucro-apos-13-anos.html>. Acesso em 28 de janeiro de 2015.

habilidades que também precisam ser aprendidas. Para Caio Túlio, a revisão dos currículos acadêmicos é fundamental, pois mudou a forma de apreensão do conhecimento. “A escola de comunicação forma para textos e não para profissionais multimídias. É preciso entender novas cadeias, novas realidades” (COSTA, 2014). De acordo com ele, é muito forte por parte dos veículos de mídia a incapacidade de entendimento de uma nova relação, inclusive sobre o posicionamento das marcas jornalísticas e modelos de negócio diante do cenário diferenciado. “A solução é entender como as lógicas funcionam, como se dão as relações” (COSTA, 2014)⁷¹.

Os autores Henry Jenkins, Joshua Green e Sam Ford mapearam no livro *Spreadable Media* (2013), traduzido para o português, em 2014, sob o título de *Cultura da Conexão*, as mudanças que ocorrem no ambiente midiático contemporâneo, espaço em que as empresas não controlam rigidamente sua distribuição de informações e muitos usuários estão diretamente envolvidos na circulação de conteúdo. Isso obriga a reflexões sobre o modelo de distribuição de mídia que parte de um conceito já consolidado para um espalhado, que instiga a todo instante novos posicionamentos de empresas e de usuários diante da rede. A lógica da comunicação digital impõe também um pensar na perspectiva da audiência, o que as tecnologias oferecem e são usadas pelas marcas, por exemplo. Para Jenkins, uma cultura da participação deve ser reconsiderada e reposicionada integrando o conceito de re-elaboração de mensagens e conteúdos, a chamada pervasividade ou mídia ubíqua, como apontado por Saad (2013).

A pervasividade ocorre quando conteúdos massivos são continuamente reposicionados ao adentrarem em diferentes comunidades de nicho na rede. Quando um material é produzido segundo a lógica “um para todos” ele resulta em inadequações às necessidades de um dado público. Diante disso, o público se vê impelido a reajustar as mensagens conforme suas necessidades informativas. (JENKINS: 2013, p.30).

A participação pervasiva pode gerar novos formatos midiáticos em uma cultura conectada, uma vez que estamos no mundo da comunicação digital, num contemporâneo onde ocorre a sobreposição de imagens, onde não há sentido no estático, onde muda-se o conceito de família e ocorre a desconstrução dos sentidos,

⁷¹ Na palestra de abertura do VII Congresso Nacional da ABCiber, realizado entre 3 e 5 de dezembro de 2014, na Escola Superior de Propaganda e Marketing – ESPM, em São Paulo, Caio Túlio ainda defende que para ele uma possível solução para os veículos de mídia é a união para trabalhar com escala. Se unir para uma network.

momento que o filósofo Peter Sloterdijk denomina de *Espumas*⁷² (2004). No âmbito deste estudo, um momento onde tem-se a necessidade de aproximar produtos jornalísticos diferenciados para um público participativo.

Em síntese, o período é de transição, de diferentes formas de apreensão do conhecimento e de apresentação da informação. Neste cenário, a narrativa pós-industrial adquire práticas diferenciadas, afinadas com o que o leitor quer e com o que as tecnologias da atualidade proporcionam. Exige um trabalho muito maior do profissional e investimentos em termos de recursos e equipe. A linguagem é imersiva, hipertextual, hipermidiática⁷³, com inúmeras fontes e correlação de dados. Para Castells, (2003), a linguagem digital é um código básico da comunicação da sociedade em rede, onde a informação é o principal fator da economia globalizada.

Jornalismo Digital de Base de Dados (JDBD) e o jornalismo de dados

Desde a década de 90 diversos estudos vem abordando as fases do jornalismo na Internet (MIELNICZUK, 2003. BARBOSA, 2007, PALÁCIOS & MACHADO, 2003 entre outros). Nomenclaturas, novas práticas, adaptação de formatos analógicos no ciberespaço. Faremos as pontuações de questões-chave, resgatando históricos para termos a percepção do cenário atual. O uso de dados no jornalismo digital tem seu marco no que a pesquisadora Suzana Barbosa (2007) denomina como o quarto estágio do jornalismo digital, marcado pela forte presença de bases tecnológicas, acessos expandidos via banda larga, proliferação de plataformas móveis, usos de algoritmos, desenvolvimento de gestão de conteúdos mais consolidados, narrativas multimídia, linguagem de programação, emprego de metadados, *data mining* (BARBOSA, 2007), um cenário que configura características específicas de um fazer jornalístico específico do ambiente online, onde as bases de dados se tornam um elemento estruturante. Alguns estudos apontam que vivenciamos já a quinta fase, a dos usos das mídias móveis para replicação de conteúdos e distribuição multiplataforma (BARBOSA & TORRES, 2013).

⁷² No livro *Esferas*, volume III, Peter Sloterdijk analisa os indícios das novas conformações da esfera em espaços sociais.

⁷³ Nas considerações da pesquisadora Lúcia Santaella “a hipermídia é uma extensão do hipertexto, pois não se limita à informação escrita, mas permite acrescentar aos textos não apenas os mais diversos grafismos (símbolos matemáticos, notações, diagramas, figuras), mas também todas as espécies de elementos audiovisuais (voz, música, sons, imagens fixas e animadas)”. (SANTAELLA, 2005, p. 24).

Entre as mais de vinte funcionalidades apontadas pelo Paradigma Jornalismo Digital de Base de Dados (BARBOSA, 2007, 2013) estão: a integração dos processos de apuração, de composição, documentação e edição dos conteúdos; a orientação e o apoio ao processo de apuração, coleta, e contextualização dos conteúdos; a regulação do sistema de categorização e qualificação *das* distintas fontes jornalísticas, indicando a relevância delas; o uso de metadados para análise de informações e extração de conhecimento, por meio de técnicas estatísticas ou métodos de visualização e exploração como o *data mining*, usos de *tagging*⁷⁴, a flexibilidade combinatória e o relacionamento entre os conteúdos.

Dentro deste paradigma encontra-se o Jornalismo Guiado por Dados, *data-driven journalism* ou como temos adotado prioritariamente neste estudo, jornalismo de dados, um “modelo que tem as bases de dados como definidoras da estrutura e da organização, bem como da composição e da apresentação dos conteúdos de natureza jornalística (...)” (BARBOSA & TORRES, 2013, p. 154). Ele “demarca a ampliação das possibilidades de emprego das bases de dados no processo de produção de conteúdos jornalísticos, no seu consumo e circulação” (ibidem, p. 154). Entre as categorias de dinamicidade, que marcam este novo padrão (em contraposição aos sites estáticos) se encontra a visualização de dados. A visualização diz respeito aos modos diferenciados para se representar informações jornalísticas, a partir da sua estruturação em base de dados.

Na década de 2000 em consequência da fase de desenvolvimento alcançada pelo jornalismo digital, pelos processos de *open data* e pelo barateamento dos processos produtivos online, o jornalismo de dados passa a envolver também o manejo de um grande volume de dados de forma automatizada e utiliza visualizações, muitas vezes interativas para encontrar notícias atrás dos dados, revelar verdades ocultas dentro de tabelas pesadas ou base de dados. Entre as novidades das redações que trabalham com este tipo de jornalismo encontra-se a incorporação de um programador nas equipes.

Uma das novidades que existe sobre este assunto é a incorporação de um programador na construção das notícias. Isto é algo completamente novo. Há cinco, seis anos, praticamente não existia. Hoje o programador se une a

⁷⁴ Tagging vem da palavra tag, em português etiqueta, é uma palavra-chave ou termo associado a uma informação que o descreve e permite uma classificação da informação baseada em palavras-chave. Disponível em: [http://pt.wikipedia.org/wiki/Tag_\(metadados\)](http://pt.wikipedia.org/wiki/Tag_(metadados)). Acesso em 28 de janeiro de 2015.

equipe de trabalho, a equipe de investigação, no desenvolvimento do que se conhece como *News Applications* ou aplicações de notícias. E existem muitos produtos em matéria de jornalismo de dados. Há artigos baseados em dados, visualizações interativas, tem aplicações de notícias, tem conjunto de dados, catálogos de dados abertos e também blogs de dados com seções dentro de alguns meios de comunicação. (CRUCIANELLI, 2014, tradução nossa)⁷⁵

O jornalismo de dados é visto como um processo de refinamento, onde dados brutos são transformados em significado. “(...) o valor para o público cresce, especialmente quando fatos complexos são moldados em histórias claras, que as pessoas podem facilmente entender (LORENZ, 2010, p. 12)⁷⁶. Os processos de refinamento dos dados representam a convergência de diversos campos significativos em um processo produtivo cada vez mais importante para as práticas jornalísticas digitais como o da pesquisa investigativa, estatística, design e programação. Para o jornalista e escritor Paul Bradshaw, especializado em jornalismo de dados, a combinação dessas habilidades resultam em histórias mais importantes e intimidadoras (BRADSHAW, 2010)⁷⁷. Para ele, o processo de produção do jornalismo de dados pode ser dividido em quatro etapas: primeiramente encontrar os dados por meio de conhecimentos, da utilização de computadores agregado a habilidades técnicas específicas; em um segundo momento a interrogação dos dados para se ter um bom entendimento do seu meio, jargões e contexto em que se aplicam; em terceiro, mesclar as informações e dar visualização aos dados por meio de infográficos e outros recursos visuais; em quarto, mesclar e apresentar os dados por meio de ferramentas próprias ou abertas.

Ao relacionar a reportagem tradicional, as narrativas com o desenvolvimento das capacidades das bases de dados, com as convergências de diferentes áreas, o jornalismo de dados representa uma forma diferenciada de jornalismo. A jornalista Sandra Crucianelli (2014) aponta que neste jornalismo o objetivo é encontrar as pérolas da notícia, contando para isso com ferramentas digitais para manusear os

⁷⁵ “Una de las novedades que hay en esta materia es la incorporación de un programador a la mesa de noticias. Esto es algo completamente nuevo. Hace 5 ó 6 años prácticamente no existía. Hoy el programador se une al equipo de trabajo, al equipo de investigación en la mesa de noticias para desarrollar lo que se conoce como *News Applications* ó *Aplicaciones de Noticias*. Y hay muchos productos en materia de periodismo de datos. Hay artículos basados en datos, visualizaciones interactivas como les decía, hay aplicaciones de noticias, hay set de datos ó catálogos de datos abiertos, pero también hay blogs de datos ó canales de datos como sección dentro de algunos medios de comunicación” (CRUCIANELLI, 2014).

⁷⁶ “(...) the value to the public grows, especially when complex facts are boiled down into a clear story that people can easily understand” (LORENZ, 2010, p. 12).

⁷⁷ “Data journalism is huge. I don't mean 'huge' as in fashionable - although it has become that in recent months - but 'huge' as in 'incomprehensibly enormous'. It represents the convergence of a number of fields which are significant in their own right - from investigative research and statistics to design and programming. The idea of combining those skills to tell important stories is powerful - but also intimidating. Who can do all that?” Disponível em: <http://www.theguardian.com/news/datablog/2010/oct/01/data-journalism-how-to-guide>. Acesso em: 10 de dezembro de 2014.

dados, muitas são ferramentas simples como o *Google Advanced Search*, a busca avançada do *Google*⁷⁸.

A pérola que buscamos e que vamos buscar no fundo do mar encontramos através de ferramentas muitas vezes conhecidas, mas pouco usadas, como a Busca Avançada do Google. Sem dúvida que esta ferramenta, através das suas planilhas, nos oferece uma versatilidade enorme na hora de buscar por formato, domínio e também por data. (CRUCIANELLI, 2014, tradução nossa)⁷⁹

Para Crucianelli, o jornalismo de base de dados reúne as técnicas do jornalismo investigativo, jornalismo em profundidade, jornalismo de precisão, jornalismo analítico e reportagem assistida por computador⁸⁰. Em um dos módulos do curso de Ferramentas de Dados para Jornalistas de Dados, Crucianelli (2014) define que:

- Jornalismo investigativo – requer três condições necessárias para que uma reportagem possa ser investigativa. Deve se tratar de um tema de relevância social, que aborde questões que alguém quer manter em segredo e, finalmente, requer o trabalho habitual do jornalista. São casos onde são investigadas corrupções, administração de dinheiro público, entre outras⁸¹.
- Jornalismo de profundidade – é o mesmo que o jornalismo investigativo, só que nestes casos não se precisa manter segredo sobre o que o jornalista investiga. Pode ser um estudo das variáveis sobre a pobreza em relação aos dados do censo⁸², por exemplo.
- Jornalismo de precisão – formulado na década de 60 pelo professor Philip Meyer,

⁷⁸ Disponível em: www.google.com.br/advanced_search.

⁷⁹ “La perla que buscamos y que vamos a bucear hacia el fondo del mar la encontramos a través de una herramienta en algunos casos muy conocida, pero poco usada, que es la Búsqueda Avanzada de Google. Sin dudas que esta herramienta a través de su plantilla nos ofrece una versatilidad enorme a la hora de buscar por formato, por dominio y también por fecha”. (CRUCIANELLI, Sandra. Curso Masivo y Abierto en Líneas - Herramientas Digitales para Periodismo de Datos. Realizado em 2014, promovido pelo *Centro de Knight para el Periodismo en las Américas*).

⁸⁰ Diante da vasta referência sobre o assunto nós não vamos nos aprofundar em cada uma destas definições. Por ora os registros ficarão por conta de Crucianelli (2014).

⁸¹ “Periodismo de Investigación (PI): Se requieren 3 condiciones necesarias para que un reportaje se precie de ser investigativo. Se trata de un tema de relevancia social a lo que se añade que alguien pretende mantener en secreto; finalmente, se requiere un trabajo adicional del periodista, ajeno a las prácticas cotidianas de cobertura informativa. Son los casos en los que se investigan asuntos relacionados con irregularidades en la administración del dinero público, como casos de corrupción, así como cuando se revelan situaciones de ineficiencia o abuso por parte de compañías de servicios públicos privatizadas o toda aquella que cause una afectación a la sociedad, como podría ser un caso de delito ecológico” (CRUCIANELLI, 2014).

⁸² “Periodismo de Profundidad (PP): Es lo mismo que el PI; solo que en este caso nadie intenta mantener en secreto lo que el periodista investiga. Como podría ser un estudio sobre las variables de pobreza sobre los datos de los censos nacionales” (CRUCIANELLI, 2014).

envolve o uso de métodos da investigação social aplicados ao jornalismo em investigações qualitativas (análises de casos, grupos focais, análise de discurso, etc) e quantitativas. Esta última inclui a descritiva (quando se consideram todos os casos: exemplo, um censo entre legisladores) e o que se denomina de investigação inferencial, ou seja, quando se trabalha com um universo e dada a dificuldade de abordar cada membro desse universo se desenha uma amostra representativa do todo, o que passa a ser um estudo projetado. Isto aplica-se a processos de votação onde há previsões, indicativos de intenção de votos⁸³.

- Jornalismo analítico – analisa uma realidade complexa, com a finalidade de alcançar a compreensão do público. Combina alguns aspectos do jornalismo investigativo. Sua missão é oferecer provas baseadas em interpretações da realidade, explicando-as de um modo compreensível. Contextualiza um assunto, descrevendo os detalhes históricos e dados estatísticos. É o resultado de uma explicação completa, destinada a dar forma a percepção do público sobre o fenômeno. Destina-se a coleta de dados díspares e a fazer conexões que não são imediatamente aparentes. Sua eficácia é na análise entre os fatos e não os próprios fatos. Usa um pilar de visualização chamado Sistema de Informação Geográfica (GIS)⁸⁴.

- Jornalismo assistido por computador ou reportagem com auxílio de computador – se define como qualquer processo que utiliza a assistência de computadores durante a relação e processamento de dados. Baseia suas práticas no uso de folhas de cálculos de Excel para a análise de informação contidas em planilhas⁸⁵.

Em relação aos produtos jornalísticos de dados, Sandra Crucianelli (2014) aponta que

⁸³ “Periodismo de Precisión (PPr): Formulado como disciplina en la década del 60 por el profesor Philip Meyer, involucra el uso de métodos de la investigación social aplicados al periodismo tales como la investigación cualitativa (análisis de casos, Grupos Focales, Análisis del discurso, etc.) y la investigación cuantitativa. Esta última incluye la descriptiva (cuando se consideran todos los casos: ejemplo, un censo entre legisladores) y lo que se denomina investigación inferencial, es decir cuando se trabaja sobre un universo y dada la dificultad para abordar a cada miembro de ese universo, se diseña una muestra representativa de ese universo, a la cual se procede a poner bajo estudio. Es el caso de las encuestas de opinión o las electorales con fines predictivos” (CRUCIANELLI, 2014).

⁸⁴ “Periodismo Analítico: Analiza una realidad compleja, con el fin de lograr la comprensión del público. Combina algunos aspectos del periodismo de investigación y la información explicativa. Su misión es ofrecer pruebas basadas en interpretaciones de la realidad, explicándolas de un modo comprensible. Contextualiza un tema de fondo, mediante la descripción de detalles históricos y datos estadísticos. El resultado es una explicación completa, destinada a dar forma a la percepción del público del fenómeno. Aspira a recoger datos díspares y hacer conexiones que no son inmediatamente evidentes, y su eficacia es a menudo en el análisis entre los hechos y no en los hechos mismos. Utiliza como pilar de la visualización los llamados SIG, Sistemas de Información Geográfica” (CRUCIANELLI, 2014).

⁸⁵ “Periodismo Asistido por Computadora: Conocido en inglés por sus siglas (CAR: Computer Assisted Reporting): se define como “cualquier proceso que utiliza la asistencia de computadoras durante la recolección y procesamiento de datos”. Ha basado su práctica en el uso de hojas de cálculo de Excel, para el análisis de información contenidas en tablas” (CRUCIANELLI, 2014).

há pelo menos seis diferentes que surgem desta prática:

- Artigos baseados em dados – geralmente são artigos curtos, construídos a partir de grande volume de dados como os que poderiam estar contidos em uma base de dados ou o conjunto de dados desenhados a partir de várias bases de dados;
- Visualizações interativas – em geral são um complemento dos artigos baseados em dados e podem ser reunidos em si mesmos sem necessidade de que exista um artigo referencial. Em alguns casos é suficiente um título, abaixo um subtítulo e uma explicação para a audiência sobre como ler os dados da visualização;
- Conjunto de dados abertos – os meios de comunicação não necessariamente se voltam para a realização de trabalhos jornalísticos baseados em dados. Alguns optam, como no caso do *La Nación*, da Argentina, por dedicar uma grande parte de seu esforço na cobertura de iniciativas governamentais *open data*. Este trabalho começa a se expandir em países onde existem leis de acesso a informação pública e onde estas leis não funcionam como deveriam. Os dados ficam confusos, em planilhas desconexas, e muitas vezes, em formatos não acessíveis via *web*⁸⁶. Nestas circunstâncias, os meios de comunicação podem dedicar esforços no trabalho com dados e certamente irão fazer grandes contribuições.
- Aplicações de notícias – conhecidas em inglês como *News Apps*. No jornalismo, com a presença de programadores informáticos, são criados aplicativos que podem permitir, por exemplo, agrupar e analisar variáveis, buscas por área geográfica, por data, nomes de empresas, etc.
- Blogs de dados – podem publicar conteúdos contendo um ou todos os itens já citados porém no formato blog;
- Canais de dados – equivalem a uma seção dentro de um meio digital. Os canais de dados incluem notas, blogs, catálogos, aplicações de notícias, etc.

⁸⁶ No próximo capítulo abordaremos as questões legislativas no cenário brasileiro.

Para o jornalista norte-americano Philip Meyer (1991), a disponibilidade de computadores nas redações e usos de bases de dados tornaram o jornalismo muito mais próximo de um método científico na sua produção. Ele defende a teoria no primeiro capítulo do livro *The New Precision Journalism*:

A disponibilidade crescente de computador fez com que grandes quantidades de informação se tornassem acessíveis aos jornalistas de um modo que antes não era possível. E, nos escritórios, o fracasso de circulação dos jornais fez com que os editores prestassem uma atenção mais sistemática ao mercado e aos fatores que motivavam os leitores a gastar tempo e dinheiro com as publicações. A noção de que o jornal é um produto e que o leitor uma criatura racional, que toma a decisão de pagar ou não por um produto, se tornou respeitável. E então as forças do mercado pressionaram o jornalismo como um todo, não somente alguns profissionais isolados, a uma postura mais científica (MEYER, 1991, tradução nossa)⁸⁷.

O aumento da disponibilidade de dados traz questões sobre como melhor apresentar essas informações para o público e como fazer isso sem que seja necessário ao jornalista um vasto conhecimento em linguagens de programação, técnicas e ciências avançadas.

Jornalismo de dados: entre o senso comum e a ciência

O jornalista e professor Philip Meyer propôs na obra *Precision Journalism* (Jornalismo de Precisão), em 1969, o uso de técnicas quantitativas das ciências sociais nas reportagens. A ideia era unir os recursos da sociologia com a rapidez de cálculo dos computadores e com isso obter abordagens mais precisas e objetivas. Neste meio entre senso comum e saber científico, o pesquisador Marcelo Träsel (2013) aponta que o senso comum é baseado na experimentação direta dos fenômenos através dos sentidos e que o saber científico é a sistematização e ordenação dos fenômenos.

⁸⁷ "The increasing availability of computers made large bodies of data available to journalists in a way that was not possible before. And in the business office, the failure of newspaper circulation to keep up with the growth in number of households made publishers pay more systematic attention to the marketplace and the factors that motivated readers to spend time and money with the publishers' products. The notion that a newspaper is a product and that a reader is a rational creature who makes a choice about whether to pay the cost of using the product became respectable. And so market forces were pushing journalism as a whole, not just a few isolated players in the field, to a more scientific stance" (MEYER, 1991). Disponível em: <http://www.unc.edu/~pmeyer/book/Chapter1.htm> . Acessado em 10/12/2015.

O “conhecimento de” (senso comum) pode ser compartilhado apenas parcialmente pela linguagem, enquanto no “conhecimento acerca de” espera-se uma universalidade (ciência). O conhecimento concretizado em forma de notícias se situaria em algum ponto do continuum entre o senso comum e a ciência. Seria um tipo de senso-comum transformado por técnicas narrativas num “conhecimento de” comunicável a toda uma sociedade (...) (TRÄSEL, 2013, p. 4).

A notícia, portanto, informa e orienta e a falta de rigor analítico muitas vezes aproximava o jornalismo mais do senso comum. O jornalismo de precisão e na atualidade o jornalismo de dados visam o outro extremo, uma aproximação às ciências através do uso de tecnologias e métodos de análise. Para Philip Meyer (1991), o jornalismo vivencia em seu cotidiano forças que visam torná-lo somente entretenimento, publicidade e relações públicas. A Reportagem Assistida por Computador (RAC), por exemplo, é o oposto:

O que os praticantes do CAR (sigla em inglês de Reportagem Assistida por Computador) estão fazendo, - conscientemente ou não, um padrão mais elevado do truthtelling. Nossa resposta para a era da informação tem sido a de aprender a gerenciar grandes massas de informações com ferramentas de análise cada vez mais poderosas, levando a uma definição mais exata da verdade. Os computadores tem sido úteis para auxiliar nisso. Mas o computador em si não é o objetivo, nem define o que estamos tentando fazer. (MEYER, 1999, p.1, tradução nossa.)⁸⁸.

A visualização de dados, tendo como um de seus elementos formadores o RAC, adquire as mesmas características de distanciamento do senso comum e aprofundamento de conteúdo, porém utilizando recursos técnicos para lidar com o grande volume e velocidade de dados aliados a estatísticas e muitas vezes aos olhares dos estudos sociais na interpretação dos fenômenos. Em resumo, há semelhanças em fazeres jornalísticos e estudos sociais, antropológicos, principalmente no que diz respeito a um jornalismo investigativo, na coleta, mapeamento e tratamento de um grande volume de dados que podem vir a representar comportamentos da sociedade. A Antropologia, por exemplo, reflete sobre o universo simbólico da ação e a estrutura

⁸⁸ “Journalism today is in a battle for survival against the forces that would merge it with entertainment, advertising, and public relations. The information age has created such a confusing buzz of voices that it tempts us all to sacrifice almost anything for attention – including truth. What the practitioners of CAR have been after – whether consciously or not – is a higher standard of truthtelling. Our response to the information age has been to learn to manage larger bodies of information with increasingly powerful analytic tools, leading to a more exact definition of truth. As it happens, a computer is helpful in doing that. But the computer itself is not the goal, nor does it define what we are trying to do” (MEYER, 1999 p. 1).

do tecido social, contribuindo não só com técnicas de análise, mas com novos olhares para o jornalisticar. Distanciando-se e aproximando-se do objeto e de visões científicas, o jornalista é um profissional que lida com todo o processo informativo: apuração, análise, síntese e transmissão de dados.

As ferramentas de computação agilizam o trabalho de fazer correlações entre dados sejam eles estruturados ou não. Trabalhando com a percepção visual contribuem com análises e expansão da capacidade de raciocínio. Por isso o jornalismo de dados tem sido visto como uma forma de reportagem com grande potencial para o cumprimento da missão de defesa do interesse público. No entanto, os dados podem ser somados, comparados e integrados entre si, complementando os grandes blocos de texto que o computador só consegue mostrar, mas não correlacionar. Esta competência fica a cargo do profissional, do jornalista, da equipe responsável pelo conteúdo jornalístico. Partindo da Antropologia Visual⁸⁹, as tecnologias atuais se constituem como um prolongamento das capacidades humanas, formas de instrumentação científica da modernidade para decifrar culturas. As tecnologias digitais tornam-se parte da memória que coletam, armazenam, organizam e comunicam uma grande quantidade de informação, em diferentes extensões e formatos. Para a pesquisadora Amber Case (2010), a tecnologia está nos evoluindo, formando uma nova versão do homo sapiens que confiam em “cérebros externos” (telefones celulares e computadores) para se comunicar. Não são extensões do eu físico, mas de um eu mental.

Você é um ciborgue toda vez que olha para a tela de um computador ou usa um celular, porque está entrando numa relação tecno-social com um pedaço de tecnologia não-humana. Nossos celulares, carros e laptops tornaram-se ciborgues porque nós os empregamos para fazer coisas que não conseguimos como simples indivíduos. Nossos corpos podem estar nos mesmos lugares, mas nossas identidades e pensamentos estão viajando pelo globo (CASE, 2010).

No livro *The Future of Antropology* (2006), a pesquisadora Sarah Pink aborda que houve uma explosão de mídias visuais nos últimos anos, uma gama de tecnologias visuais e digitais que transformaram a investigação e análise visual. Ela aponta como resultado uma abordagem interdisciplinar de grande potencial dentro e fora da

⁸⁹ A antropologia visual é um ramo da antropologia cultural, aplicada ao estudo e produção de imagens, nas áreas da fotografia, do cinema ou, desde os meados dos anos 1990, nos novos “*media*” utilizados em etnografia. Fonte Wikipédia. Disponível em: http://pt.wikipedia.org/wiki/Antropologia_visual. Acesso em 10 de junho de 2014.

academia. Argumenta que esse potencial pode ser aproveitado através das concepções da Antropologia Visual em contextos mais amplos, incluindo:

- O aumento do uso de métodos de pesquisa visuais através das ciências sociais e humanas;
- O crescimento da popularidade do visual como metodologia e objeto de análise dentro da antropologia e da antropologia aplicada
- O crescente interesse na antropologia dos sentidos e antropologia da mídia
- O desenvolvimento de novas tecnologias visuais que permitem aos antropólogos trabalhar com novas formas. (PINK, 2006, tradução nossa).

O futuro da visualização dos dados depende em parte, de como os seus praticantes vão se envolver com as novas mídias visuais e digitais, com os avanços tecnológicos. De acordo com Sarah Pink tornam-se essenciais para os pesquisadores três fases de uma análise visual: produção, conteúdo e contexto de consumo. Onde são revelados vários processos de produção e relações que envolvem diferentes configurações de indivíduos e instituições. Conteúdo e design também são elementos importantes na forma de apresentação e representação da comunicação hipermídia, relacionando principalmente teoria e experiência numa estratégia para se alcançar os objetivos. O consumo é de igual importância, uma vez que é fundamental pensar sobre como o trabalho foi recebido e qual o seu impacto com o público. Ao combinar imagem, som e escrita com a relação de sentidos é preciso se envolver com as epistemologias, de como essa experiência é construída; saber a melhor forma de explorar e de analisar os diferentes aspectos da experiência humana através de diferentes meios de comunicação. É preciso repensar o visual em múltiplos níveis de engajamento e diversos públicos, atentando para os desafios que surgem a partir da experiência e da representação.

Qual é a visão e como ela opera tanto como uma experiência corporal como uma categoria cultural em diferentes contextos sociais e culturais? Além disso, como são as relações especificamente culturais entre visão e outras categorias sensoriais construídas e como são relacionadas as convenções culturais e as intervenções humanas para quebrá-las ou modificá-las? Em segundo, tal repensar exige interrogar os nossos métodos e meios de comunicação visual em pesquisa, seus usos para examinar a relação entre (áudio) visual e outros meios de comunicação (como a escrita, desenho, mídia digital, arte de instalação, performance). Como tal, a Antropologia Visual pode ser redefinida não apenas como a antropologia do visual e do uso de métodos visuais na pesquisa e na representação, mas como a antropologia da relação entre os elementos visuais e outras formas de cultura, a sociedade, a prática e a experiência e a prática metodológica

combinando meios de comunicação visual e outras na produção e representação do conhecimento antropológico (PINK, 2006, p. 143, tradução nossa)⁹⁰.

Sociedade complexa, recursos tecnológicos nunca vistos e vividos. Para o fazer jornalismo de dados, para a compreensão do social, são necessários novos olhares, uso de métodos de pesquisa e diferentes formas de apresentação dos dados que facilitem o conhecimento por parte do público.

Considerações

Este primeiro capítulo teve por objetivo fazer uma contextualização do jornalismo contemporâneo, onde ele se encontra, assim como onde estão os estudos do jornalismo digital no que diz respeito ao grande volume de dados. Buscamos de forma ampla, a partir da Comunicação, situar o contexto do objeto de estudo proposto inserindo-o num cenário com mais rigor científico, com demandas de contextualização aprofundada, em uma linguagem imersiva, dinâmica, com a proximidade de outras áreas de pensamento e investigação.

Pontuamos, ao longo do capítulo, a distinção entre dado e informação e entre visualização de dados e visualização da informação. Deixamos claro nosso desafio nesta pesquisa de focar não em um resultado final, mas na compreensão sobre como os processos se estabelecem no âmbito deste manuseio de dados e metadados. Para nós a visualização de dados é um processo não-linear, trabalhado no jornalismo de dados que se encontra em área fronteira de estudo, mas possuindo características próprias que o distingue de outras práticas. Ele é um jornalismo visual, investigativo, de precisão, analítico, de profundidade, que envolve as técnicas do RAC e que emerge a partir da congruências e manuseio de grandes volumes de dados. Nos

⁹⁰ “What is vision and how does it operate as both an embodied experience and a cultural category in different social and cultural contexts? Moreover, how are culturally specific relationships between vision and other sensory categories constructed and how are these related to cultural conventions and human agency to break or change them? Second, such a rethinking requires that we interrogate our uses of visual methods and media in research, to examine the relationship between (audio) visual and other media (such as writing, drawing, digital media, installation art, performance). As such, visual anthropology might be redefined as not simply the anthropology of the visual and the use of visual methods in research and representation, but as the anthropology of the relationship between the visual and other elements of culture, society, practice and experience and the methodological practice of combining visual and other media in the production and representation of anthropological knowledge” (PINK, 2006, p. 143).

concentramos, prioritariamente, no fator tecnológico como um dos pré-requisitos para estudar as peças jornalísticas que serão exploradas no capítulo V e que acreditamos que vão nos ajudar a sustentar a resposta da pergunta-problema desta pesquisa: o que é a visualização de dados no jornalismo? Como ela se insere na prática e na narrativa do jornalismo de dados proporcionando diferenciais relevantes?

Como chegamos a abordagem da visualização de dados no sentido trabalhado nesta pesquisa? De forma visual, nosso esquema poderia ser descrito desta forma:

Figura I. 5 – Esquema conceitual da visualização de dados nesta pesquisa

Fonte: Mayanna Estevanim, 2016.

Dentro de um cenário Big Data, o jornalismo de dados, que adota elementos do jornalismo investigativo, visual, de precisão, analítico, da Reportagem Assistida por Computador e de profundidade, se relaciona com outras áreas para lidar com o grande volume, velocidade e variedade de dados. Estas áreas são a Comunicação, a Computação, o Designer, e a Estatística (prioritariamente). Entre seus recursos, o jornalismo de dados adota a visualização de dados como formato para interpretar as informações (visualizar as informações) e apresentá-las em narrativas inovadoras, seja a partir de infográficos, mapas, geolocalização ou ferramentas interativas.

Nos próximos capítulos, partimos do olhar de que a visualização de dados é um formato que recorre à utilização de *softwares* de processamento de dados que estão envolvidos nas camadas de dados e metadados do sistema narrativo, traduzindo complexidades de dados e suas correlações através de modelos visuais. As bases de dados não necessariamente aparecerão ao usuário/leitor, uma vez que podem servir apenas como fonte para a produção de conteúdos jornalísticos, além de apresentações infográficas, mapas, entre outros recursos visuais. Para nós a visualização aqui é um meio de transformar dados em informação semântica, criando ou não ferramentas para que qualquer pessoa faça a leitura e interpretação de um processo informativo.

II. A ENTREVISTA AOS DADOS

Resumo: O segundo capítulo tem como meta exemplificar o jornalismo de dados e a visualização para, a partir da sua presença no cotidiano, discutir as questões teóricas e apontar os caminhos para construções noticiosas que partem de base de dados. Tendo como o foco no cenário brasileiro, apontar a influência das questões políticas e de leis de transparência que interferem na produção e disseminação da usabilidade na *web* e da prática do jornalismo de dados no país.

Palavras-chave: base de dados, visualização de dados, jornalismo de dados brasileiro

Saber codificar e decodificar tecnocódigos, para contar histórias, deverá ser muito em breve uma necessidade básica de todos (Ligia Maria Prezzi Lemos, 2013, em seminário sobre a obra de Vilém Flusser)⁹¹.

A avalanche de informações

No ano de 1854 uma epidemia de cólera ocorreu na cidade de Londres, Reino Unido (UK) contabilizando mais de 500 vítimas fatais num período de apenas 10 dias. O bairro de *Soho* reuniu um grande número de moradores contaminados. O principal foco de transmissão da doença foi uma bomba d'água onde os moradores e transeuntes enchiam seus vasilhames. O médico Jonh Snow se destacou na investigação dessa epidemia ao reunir os registros de cada vítima e onde elas residiam, quais trajetos faziam diariamente e representar isso através de um mapa com o objetivo de identificar padrões de contágio. Desta forma ele conseguiu mostrar indícios fortes de que a doença não seria transmitida pelo ar, mas sim pela água e especificamente que aquele surto partia de um poço.

⁹¹ Realizado dia 11 de junho de 2013, na disciplina Comunicação e redes na era digital: pensamento contemporâneo e tendências tecnológicas, ministrada pela professora doutora Elizabeth Saad, na Escola de Comunicação e Artes – USP.

Figura II. 6 – Mapa de Snow, 1854.

Fonte: JOHNSON, 2008, p. 117

Snow realizou um intenso trabalho de investigação de campo, foi nas áreas onde ocorriam a maior parte das mortes; fez diversas entrevistas com moradores da localidade e de áreas próximas, observou a relação das mortes que ocorriam em áreas mais afastadas e ao visualizar as conexões descobriu que a maioria das pessoas se concentravam em torno de um poço de água localizado na *Broad Street*. Com este mapa, Snow foi considerado um dos pioneiros nos estudos sobre epidemiologia.

Em seu formato original, apresentado ao público em um encontro da Sociedade Epidemiológica (...) cada morte foi representada por um grosso traço preto, o que proporcionava um vívido destaque às casas que sofreram uma quantidade significativa de perdas, e o excesso de detalhes foi eliminado, preservando-se unicamente o traçado básico das ruas e os símbolos que representavam as treze bombas d'água que abasteciam a

vasta área do *Soho*. O impacto visual do mapa era impressionante. (...) Era possível observar que onze bombas d'água não apresentavam casos de cólera nas proximidades (...) quando se enfatizava a imagem da bomba de Soho, o mapa ganhava uma súbita clareza. O cólera não se estendia difusamente sobre o bairro. Irradiava-se, na verdade, a partir de um único ponto. (JOHNSON, 2008, p. 77).

A criação do mapa filtrando somente as variáveis que importavam para combater o problema sintetizou a complexidade presente no cenário e favoreceu o conhecimento através da representação visual. “O que fortaleceu esse modelo de visualização não foi exatamente a técnica de mapeamento empregada, mas sim a ciência por trás da representação, ou seja, a intenção explícita de amplificar a capacidade de cognição do observador” (RIBEIRO, 2009, p. 40). O trabalho de Jonh Snow foi árduo e revelador. De 1854 até os dias atuais foram inúmeras as transformações, sejam nas formas de investigação de epidemiologias, no armazenamento dos dados ou na digitalização dos mesmos, o que possibilita diferentes olhares sobre a sociedade em seus mais diversos aspectos. Falamos aqui de processamentos e correlações de bancos de dados que são humanamente impossíveis de serem mensurados de forma manual, principalmente se atentarmos para o fato de que estamos na casa dos yottabytes⁹² de dados.

Voltando o nosso olhar para o jornalismo, entre os exemplos que associam a comunicação em rede (diante de um grande volume de dados) e o jornalismo investigativo no Brasil, se encontra o SwissLeaks⁹³, uma investigação sobre o gigantesco esquema de evasão fiscal onde o banco multinacional britânico HSBC tinha conhecimento e facilitava as ações através de uma de suas subsidiárias na suíça, o *HSBC Private Bank Suisse*. Nestas ações, o banco teria obtido lucros ao abrigar dinheiro de sonegadores fiscais e outros transgressores de leis. O trabalho de investigação foi coordenado pelo Consórcio Internacional de Jornalistas Investigativos – ICIJ (sigla em inglês de International Consortium of Investigative Journalists⁹⁴) – e envolveu mais de 160 jornalistas em 55 países incluindo França, Estados Unidos, Suíça e Brasil. Os jornalistas revelaram que 180,6 bilhões de euros foram movimentados em contas mantidas no HSBC, em Genebra, oriundas de todo o mundo, entre novembro de 2006 e março de 2007. Estas contas representavam

⁹² Um yottabyte (10^{24}) equivale a 250 trilhões de DVDs de dados, mas já se falam nos BrontoBytes de dados (10^{27}) no universo digital.

⁹³ Entre as fontes, informações iniciais podem ser obtidas em: https://pt.wikipedia.org/wiki/Swiss_Leaks; <http://fernandorodrigues.blogosfera.uol.com.br/2015/03/12/leia-tudo-sobre-o-caso-swissleaks-hsbc-no-brasil/>.

⁹⁴ <https://www.icij.org/index.html>

mais de 100.000 clientes e 20.000 empresas offshore⁹⁵. Os registros apontavam que do montante, 8.667 eram correntistas brasileiros, que totalizavam depósitos em cerca de US\$ 7 bilhões. O Brasil foi o 4º país em número de clientes envolvidos.

Figura II. 7 – Globo.com, matéria A distribuição dos clientes e valores do "Swissleaks", 2015.

Fonte: Jornal O Globo, 2015.

Um dos infográficos interativos apresentados na imprensa⁹⁶ mostrou a distribuição dos clientes por países e valores não declarados. O resultado foi obtido através do cruzamento de bancos de dados originados de arquivos roubados do *HSBC Private Bank* por um ex-funcionário, o engenheiro de *software* Hervé Falciani. Estes mesmos arquivos foram entregues às autoridades da França no final de 2008 e também para

⁹⁵ Uma empresa offshore é uma empresa registrada em um país que oferece um sistema fiscal particularmente vantajoso ou até mesmo inexistente, desde que a atividade não seja realizada no território de registro da empresa. As empresas offshore têm três características principais: primeiro, elas devem ser registradas na jurisdição escolhida. Em segundo lugar, o cliente deve ser domiciliado fora da jurisdição de registro. Em terceiro lugar, a empresa deve realizar todos os seus negócios fora da sua jurisdição de registro, se não seria considerado como uma empresa "residente" e portanto, tributados como tal. A maior parte das pessoas associa o termo "empresa offshore" como sendo uma forma de otimizar o sistema tributário da empresa. Fonte: <http://pt.sfm-offshore.com/faq-empresaooffshore.html#S0>. Acesso em 15 de dezembro de 2015.

⁹⁶ Disponível em: <http://infograficos.oglobo.globo.com/brasil/a-distribuicao-dos-clientes-e-valores-do-swissleaks.html>. Acesso em 11 de dezembro de 2015.

diversas autoridades de outros países, entre elas a brasileira (que em um primeiro momento negligenciou a informação). Nosso objetivo aqui é aprofundar um pouco mais o olhar a este exemplo para perceber a dimensão do cenário de dados investigado, o impacto do manuseio dessas informações para o jornalismo de dados e para a sociedade.

O passo-à-passo

A análise dos cerca de 60 mil arquivos sobre as contas bancárias passou por diversas etapas e contou com a experiência de quem já vinha lidando com denúncias e um grande volume de dados. A jornalista americana Jessica Weiss (2015), que faz parte da Rede de Jornalistas Internacionais IJNet conversou com a vice-diretora do Consórcio Internacional de Jornalistas Investigativos, Marina Walker Guevara e com a jornalista de dados Mar Cabra, da mesma instituição, sobre como foi este trabalho em escala global⁹⁷. Primeiramente os dados foram recebidos pelo ICIJ que teve o cuidado de examiná-los e fazer listas regionais e nacionais de nomes e clientes para tornar mais fácil o trabalho para os repórteres locais. Somente alguns meses depois foi convocada uma reunião presencial, de um dia, no jornal *Le Monde*, em Paris, com todos os repórteres que faziam parte do projeto.

O programador Rigoberto Carvajal esteve a frente dos dados em um primeiro momento e foi quem escreveu um programa para extrair informações simples dos arquivos para serem convertidas em um banco de dados. Estas informações foram repassadas aos repórteres para que tivessem algo palpável para iniciar os trabalhos. A comunicação entre os participantes ocorreu a partir de um *software* de rede social de código aberto chamado *Oxwall*, que foi personalizado pelo ICIJ e renomeado de *Voyager*. De acordo com Mar Cabra, para os dados foram utilizadas duas plataformas de código aberto.

Para pesquisas de documento usamos a plataforma de busca Solr de código aberto, o que nos permitiu categorizar e fazer o upload dos vazamentos em uma plataforma segura, e então permitiu que os repórteres pesquisassem os vazamentos usando palavras-chave, país, ano e outras variáveis. Além do

⁹⁷ Disponível em: <https://ijnet.org/pt-br/blog/como-o-icij-fez-uma-investiga%C3%A7%C3%A3o-transfronteiri%C3%A7a-colaborativa-e-em-grande-escala>. Acesso em 05 de janeiro de 2016.

Solr, também usamos Blacklight, que é uma interface de usuário que tornou a ferramenta mais fácil de usar (CABRA apud WEISS, 2015).

Além dos desafios operacionais para a construção do banco de dados próprio, os desafios jurídicos do projeto estiveram presentes durante todo o processo, uma vez que estavam lidando com questões delicadas, onde não podiam haver erros neste banco de dados. O manuseio com os dados, a chamada limpeza que envolve o agrupamento por familiaridade e exclusões, trouxe para os jornalistas envolvidos na investigação a percepção de que as conexões emergentes entre os dados e as pessoas eram a parte mais importante. Por isso foi criada uma ferramenta de código aberto, denominada *Linkurious*, que permitiu a conexão direta das pessoas aos dados. Foi necessária a elaboração de um manual sobre como usar as plataformas, assim como treinamentos online para mostrar aos jornalistas como ler os arquivos. Para as comunicações entre os envolvidos foi utilizado um serviço de e-mail chamado *Hushmail*.

De acordo com Marina Walker Guevara as conquistas foram notáveis principalmente pela união de esforços entre redações que passam por crises econômicas. “Nenhuma redação até agora - exceto talvez o New York Times e alguns outros - têm a capacidade de fazer algo grande como este em uma escala global” (GUEVARA apud WEISS, 2015).

Compartilhar não é um instinto de jornalistas -- simplesmente não é a maneira que nós tradicionalmente trabalhamos. Mas nós incentivamos isso. Estamos tentando mudar a forma como funciona o jornalismo investigativo, porque achamos que isso é como ele pode e deve ser no futuro. [Profissionais de] negócios, a polícia, criminosos -- já estão todos trabalhando de forma colaborativa. Como jornalistas, nós somos os únicos que estamos atrás (GUEVARA apud WEISS, 2015).

Em fevereiro de 2015, foi divulgada na página do ICIJ a informação sobre as contas na Suíça, com o título de *Swiss Leaks: Murky Cash Sheltered by Bank Secrecy*⁹⁸. No Brasil, o UOL iniciou a participação em fevereiro deste mesmo ano em seguida o Jornal O Globo se integrou a iniciativa. O jornalista Fernando Rodrigues, do UOL e o

⁹⁸ Swiss Leaks: dinheiro escuso protegido pelo sigilo bancário (tradução nossa).

jornalista Chico Otávio, do Globo, estiveram envolvidos na investigação que no total divulgou 500 nomes de brasileiros. Em palestra realizada no 10º Congresso Internacional de Jornalismo Investigativo da Abraji, Associação Brasileira de Jornalismo Investigativo⁹⁹, em 2015, os dois deram detalhes sobre a investigação que esteve durante 6 meses concentrada somente com programadores no processo de depuração das contas e formação de um banco de dados nacional próprio com o uso da ferramenta *Linkurious*. A partir desse processo foi possível perceber a teia de relações, transferências de dinheiro, transferências de quem para quem e foram adotados os critérios de divulgação e escolha dos nomes que estariam nas matérias nacionais.

Para a vertente brasileira do caso foi montada uma editoria temporária à parte nos veículos envolvidos. Primeiramente o objetivo era entender quais os limites legais para ter contas fora do país. Foram feitas diversas perguntas aos dados e a diferentes fontes até que fosse formado um banco de dados interno. Foram necessárias criações de contas, provedores específicos para se ter mais segurança contra invasões de e-mails e cibercrimes¹⁰⁰ no período de investigação. A checagem das informações foi diária e aos poucos foram elaborados os critérios de seleção. A mineração envolvia a leitura de nomes que estavam incompletos e por isso foi preciso o cruzamento com datas de aniversário e outras correlações para conseguir chegar aos nomes que foram escolhidos pelos critérios de relevância jornalística combinada com interesse público. Em suma, pessoas que tinham exposição pública, que verbalizavam opinião, influenciavam a sociedade e que viviam dessa exposição pública.

Primeiramente foi feito o rastreamento dos correntistas brasileiros (priorizando os que tinham registros de fraudes públicas). Aqueles que haviam declarado à Receita Federal e que o Banco Central tinha conhecimento sobre as contas no exterior saíram desta lista. No entanto, menos de 10%, dos mais de 8 mil nomes, tinham estes esclarecimentos prestados. Todos os nomes reunidos nesta lista de 500 pessoas foram contactados e informados sobre as divulgações que iriam acontecer e foram

⁹⁹ Palestra proferida em 2 de julho de 2015, na Universidade Anhembi Morumbi, na Vila Olímpia, em São Paulo.

¹⁰⁰ Cibercrime (*Cybercrime* em inglês) é a palavra dada a uma prática que consiste em fraudar a segurança de computadores ou redes empresariais. Este crime pode ser promovido de diversas maneiras: disseminação de vírus que coletam e-mails para venda de *mailing*; distribuição material pornográfico (em especial infantil); fraudes bancárias; violação de propriedade intelectual e direitos conexos ou mera invasão de sites para deixar mensagens difamatórias como forma de insulto a outras pessoas. Disponível em: <https://pt.wikipedia.org/wiki/Cibercrime>. Acesso em 25 de abril de 2016.

convidados a prestar esclarecimentos. Diversos advogados surgiram e ligações foram feitas diretamente para a diretoria dos veículos solicitando esclarecimentos.

De posse dos nomes, a divulgação aconteceu por temas, entre eles de pessoas envolvidas em escolas de samba, no mundo do esporte, artistas, empresários. A divulgação foi coletiva, pois havia um acordo assinado entre os jornalistas dos diversos países envolvidos no consórcio. O banco de dados constituído pelas equipes de mídia não foi disponibilizado para os órgãos públicos, somente partes dele para esclarecimentos, uma vez que a preocupação jornalística também incidiu sobre aqueles que estavam na legalidade e, portanto, tinham direito a privacidade. Advogados criminalistas alegaram origem ilegal da prova e os desdobramentos sobre o caso correm na justiça.

A investigação SwissLeaks foi demorada, envolveu um grande número de pessoas e teve como suporte o conhecimento do jornalismo investigativo aliado aos recursos tecnológicos e análises estatísticas para lidar com o volume de dados na contemporaneidade. Como resultando, teve a construção de bancos de dados próprios de cada país envolvido e inúmeras repercussões nos mais diferentes formatos. Com a digitalização das informações e com a possibilidade de manusear estes conjuntos de dados, o trabalho jornalístico foi dimensionado, ganhou escala e profundidade em relação aos dados. O que está de acordo com a nossa visão de que as maneiras e finalidades de utilização da prática de visualização de dados são diferenciadas e envolvem aspectos culturais, de acessibilidade de regulamentações. Por isso, vamos nos aprofundar no cenário nacional como meio de entender particularidades do processo.

Visualização de dados no jornalismo digital brasileiro

Como apontado no primeiro capítulo, para nós a visualização de dados é “o aproveitamento do agenciamento entre as camadas de dados e metadados no âmbito do sistema narrativo, o qual privilegia a apresentação visual relevante e de impacto” (BERTOCCHI, 2013, p. 180). É um formato que recorre à utilização de *softwares* de

processamento de dados que não necessariamente aparecem ao usuário/leitor, uma vez que os resultados finais podem ser construções narrativas a partir dos dados obtidos, como infografias, mapas, entre outros¹⁰¹. Variáveis em relação as equipes envolvidas, desenvolvimento de *softwares* próprios, recursos em geral e principalmente disponibilidade de tempo de execução e entrega dos conteúdos aos leitores. Como o jornalismo nacional vem se apropriando da visualização de dados e das práticas do jornalismo de dados? A partir de uma observação não-participante desde janeiro de 2014, em decorrência do ingresso do mestrado em Ciências da Comunicação na Universidade de São Paulo, nós distinguimos, em um primeiro momento, três principais áreas que trabalham a visualização de dados em âmbito nacional: os veículos tradicionais de mídia, as iniciativas independentes e os laboratórios acadêmicos.

Entre os exemplos nacionais, o Blog Estadão Dados¹⁰², do Jornal O Estado de S. Paulo, é um veículo tradicional de mídia que possui uma equipe própria para trabalhar com jornalismo de dados. O blog apresenta tanto reportagens, quanto projetos em visualização de dados e criou a ferramenta chamada Basômetro¹⁰³ que permite medir o desempenho dos governos e dos deputados brasileiros. O resultado final apresentado ao leitor neste projeto foi uma infografia interativa¹⁰⁴ que utiliza como recursos visuais uma timeline com bolas, linhas e porcentagem para medir a taxa de governismo (medição do apoio ou não dos deputados ao governo) do parlamentar a partir da recuperação interna das bases de dados do sistema construído.

Voltaremos aos exemplos do Estadão Dados no capítulo 5, inclusive aprofundando questões específicas do Basômetro.

Figura II. 8 – Estadodados.com, projeto Basômetro, 2008.

¹⁰¹ O foco deste artigo é o cenário nacional, mas as publicações em âmbito internacional são avançadas e inovadoras no jornalismo digital. O jornalista americano Josh Stern publicou um post no seu blog intitulado Groundswell com as reportagens que foram destaque no jornalismo em 2014. Elas utilizaram storytelling, recursos de áudio, vídeo, satélites, sensores, imersão. O post The Best Online Storytelling and Journalism of 2014, está disponível em <http://stearns.wordpress.com/2015/01/05/the-best-online-storytelling-and-journalism-of-2014/>. Acesso em 12 de janeiro de 2015.

¹⁰² <http://blog.estadaodados.com>

¹⁰³ <http://estadaodados.com/basometro/>

¹⁰⁴ A palavra infografia vem do inglês informational graphics e consiste em utilizar ferramentas visuais como mapas, tabelas, desenhos, ilustrações, legendas, combinadas com texto no intuito de transmitir ao leitor informações de fácil compreensão (CAIRO, 2004, DE PABLOS, 1999),.

Fonte: Jornal O Estado de S. Paulo, blog Estadão Dados, 2008.

Outro exemplo brasileiro é o Hacks Hackers¹⁰⁵ SP, um grupo independente que une pessoas de diversas áreas em ações colaborativas para explorar tecnologias de filtragem e visualização de informações. O grupo busca diferentes maneiras de contar histórias no jornalismo. Eles realizam as denominadas “hackatonas”, eventos que reúnem programadores, designers, jornalistas e pesquisadores em torno de dados públicos, sobre ideias de visualização da informação e monitoramento de propostas já iniciadas. Cada evento conta com uma temática e parcerias, a exemplo do *Hack d'Água'São Paulo: dados e sensores para informar sobre a crise hídrica*¹⁰⁶, que conta com o Instituto Socioambiental¹⁰⁷ e a equipe do Lab (InfoAmazônia¹⁰⁸ & Mapas Coletivos¹⁰⁹) como facilitadores, apresentando trabalhos já desenvolvidos tanto na visualização de dados, quanto em sensores para a medição da qualidade da água.

Figura II. 9 – Encontro Hack d'Água'São Paulo, 2014

¹⁰⁵ <http://hackshackers.com>. Outras informações também estão disponíveis em: <http://www.meetup.com/Hacks-Hackers-Sao-Paulo/>.

¹⁰⁶ <http://www.meetup.com/Hacks-Hackers-Sao-Paulo/events/206050432/>

¹⁰⁷ <https://www.socioambiental.org/pt-br>

¹⁰⁸ <http://infoamazonia.org/pt/#!/map=49&loc=-5.659718554577286,-61.25976562499999,5>

¹⁰⁹ <http://www.mapascoletivos.com.br/>

Fonte: Hacks Hackers, 2014¹¹⁰.

O próximo exemplo atrela o jornalismo, a visualização de dados e os estudos de análise de redes. Ele vem do Laboratório de Estudos sobre Internet e Cultura do Departamento de Comunicação da Universidade Federal do Espírito Santo (LABIC/UFES)¹¹¹. Como a própria descrição da sua página na *web* diz, é um laboratório que tem como proposta a realização experimental de produtos digitais e a promoção de pesquisas e atividades de extensão sobre o impacto da cultura digital nos processos e práticas de comunicação contemporânea. O grupo é formado por jornalistas, programadores, universitários, entre outros. A cartografia *A longa jornada do #MarcoCivil: da letargia do Parlamento ao efeito Snowden*, foi um trabalho que utilizou o *software* livre *Gephi*¹¹², uma plataforma interativa de visualização e exploração de dados em grafos¹¹³. A análise propôs a trajetória das discussões sobre o #MarcoCivil¹¹⁴ nas redes sociais.

¹¹⁰ <http://www.meetup.com/Hacks-Hackers-Sao-Paulo/events/206050432/>

¹¹¹ <http://www.labic.net>

¹¹² O *Gephi* pode ser baixado gratuitamente no <https://gephi.org>.

¹¹³ “Um grafo pode ser compreendido como a representação gráfica de uma rede se valendo de elementos como nós (pontos de conexão) e arestas (conexões entre os nós)”. (DI FELICE, Massimo, TORRES, Julliana C. e YANAZE, Leandro K. H. Redes

Figura II. 10 – Labic.net, cartografia A longa jornada do #MarcoCivil, 2012

Fonte: Labic/UFES, 2012

A equipe do laboratório registrou, em 2012, 22 mil tweets somente com a hashtag #MarcoCivil. Destes, foram separadas as mensagens que viralizaram na rede, um total de 5 mil retweets. Com a leitura dos retweets, as mensagens mais viralizadas, por diferentes atores, se concentrava em controvérsias em torno de: questões políticas, efetivação da lei, neutralidade da rede, liberdade de expressão, direito autoral, direito do consumidor, cibercrimes e privacidade. A partir disso, com estudos técnicos sobre atribuições de cores e arestas, foram criados gráficos sobre a diversidade de controvérsias, as oscilações diante dos novos acontecimentos, como adiamento da votação. Foi em 2013, que de acordo com o estudo, a rede deixou de ser de ativistas, políticos e especialistas e se popularizou. Esta popularização foi ajudada pela entrada em cena da rede contrária #NãoVaiTerMarcoCivil, pela realização das jornadas de

digitais de sustentabilidade: As interações com o meio ambiente na era da informação. São Paulo: Annablume. 2012).

¹¹⁴ O Marco Civil da Internet é a lei que regula o uso da Internet no Brasil (previsão de princípios, garantias, direitos e deveres para quem usa a rede, além dos usos por parte do estado estão entre a abrangência da lei). O projeto surgiu em 2009, mas foi somente no primeiro semestre de 2014 que ele foi sancionado. Informações sucintas podem ser encontradas em http://pt.wikipedia.org/wiki/Marco_Civil_da_Internet. Acesso em 07 de janeiro de 2014. Aprofundaremos as questões sobre o Marco Civil no decorrer deste estudo.

junho¹¹⁵ e pela divulgação da espionagem eletrônica que o governo dos EUA foi denunciado por praticar¹¹⁶. Em todo este percurso é possível ver como a rede reagiu a cada fato novo.

A partir dos exemplos apresentados neste capítulo é possível perceber que há diferentes apropriações, técnicas, métodos de trabalho e análise diante de um universo de opções de designs e fontes de dados. Entre os pontos em comum temos o trabalho em equipe que envolve pessoas de diversas áreas, mas que nas redações jornalísticas se concentram principalmente entre jornalistas, estatísticos, programadores e designers. É possível perceber que dentro das variáveis, **há características inerentes ao jornalismo de dados. De imediato as características que envolvem: a formação das equipes, a temporalidade na produção dos conteúdos e as distintas fontes, que acabam por envolver também a expertise dos profissionais envolvidos na coleta dos dados.**

Características inerentes

Equipes

O Estadão Dados tem um núcleo hoje formado por quatro profissionais: um coordenador, dois jornalistas e um estatístico. Na hackatona da água estiveram presentes 43 profissionais das mais diferentes áreas. O Labic contava, em 2015, com o envolvimento de 20 profissionais, num trabalho multidisciplinar envolvendo pessoas dos departamentos de Ciências da Comunicação, Computação, Design, Matemática e Antropologia para fazer a análise das redes sociais e navegar desde a cibercultura, passando pelos movimentos de rede até a estética contemporânea. No caso do SwissLeaks houve o agrupamento de profissionais de diferentes empresas de comunicação em um consórcio. Falamos, portanto, de um trabalho em equipe, com pessoas com diferentes competências.

¹¹⁵ Protestos realizados no Brasil em 2013, também conhecidos como Manifestações dos 20 centavos, Manifestações de Junho ou Jornadas de Junho. As manifestações populares aconteceram em várias regiões do Brasil e foram iniciadas para contestar o aumento da tarifa do transporte público.

¹¹⁶ Edward Joseph Snowden é um analista de sistemas que trabalhou na Central Intelligence Agency (CIA) dos Estados Unidos e também na Agência de Segurança Nacional (ANS) Americana, que denunciou o sistema global de vigilância feito pelo governo americano.

Tempo de produção

Não há um tempo definido para a consolidação do banco de dados, correlações e produção das peças jornalísticas. Como em boa parte das vezes, o trabalho se assemelha ao do jornalismo investigativo podendo demorar dias ou meses. A lógica é completamente diferente dos processos das redações tradicionais de jornalismo que trabalham a atualização cotidianamente.

Alguns veículos de comunicação tem modificado sua rotina mesmo não trabalhando diretamente com jornalismo de dados e visualização de dados. No fim de março de 2016, por exemplo, o jornal britânico Times¹¹⁷ anunciou que não publicaria mais notícias em tempo real tanto no site, quanto no aplicativo. As atualizações, com exceção de notícias urgentes, acontecem agora duas ou três vezes ao dia. O objetivo do jornal é dar mais profundidade a assuntos considerados relevantes¹¹⁸. No próximo capítulo abordaremos questões referentes ao valor-notícia e a inovação, que talvez possam ser justificativas para mudanças como esta.

Fontes dos dados

A escolha das fontes, a seleção de informações e o formato da narrativa são práticas comuns no jornalismo independentemente da digitalização. Na contemporaneidade, os dados são oriundos de políticas de transparências, dados divulgados pelos governos e instituições sociais nas mais diferentes extensões. Além dos dados das redes sociais, coletados em tempo real. Ou, no caso do Labic, a partir de uma ferramenta própria elaborada para coletar dados da rede feita posteriormente aos acontecimentos, principalmente no Twitter.

Em 2014, em artigo apresentado no IV Congresso Internacional de Jornalismo – Obciber, em Portugal¹¹⁹, foram pesquisados os principais jornais das capitais do Brasil¹²⁰ como forma de contextualização do uso de visualização de dados e

¹¹⁷ <http://www.thetimes.co.uk/>

¹¹⁸ Disponível em: <https://labmidiaombr.wordpress.com/2016/05/04/breaking-news-nao-pera-ai/>. Acesso em 5 de maio de 2016.

¹¹⁹ RAMOS, Daniela; SPINELLI, Egle; ESTEVANIM, Mayanna. Formatos digitais e qualidade no jornalismo brasileiro em base de dados, 2014. Disponível em: <https://cobciber4.files.wordpress.com/2015/04/cobciber4atas.pdf>. Acesso e, 2 de junho de 2015.

¹²⁰ Foram buscados os jornais catalogados no site: <http://www.guiademidia.com.br/jornais-brasileiros.htm>.

infografia no jornalismo online brasileiro. A observação aconteceu nos dias 15 e 16 de setembro de 2014 diretamente no endereço eletrônico de veículos de comunicação. Foram observados sites e portais de notícia que apresentaram conteúdo produzido especialmente para o formato online. Foram selecionados quatro jornais, aleatoriamente, de cada capital, totalizando, na primeira observação, 108 veículos. Em um segundo momento foram buscadas seções “dados” ou “infografia” nas páginas dos veículos. Dos 108 observados em apenas cinco veículos de cinco Estados (Goiás, Rio Grande do Sul, Ceará, São Paulo e Rio de Janeiro) foram encontrados blogs ou abas sobre o assunto (RAMOS et al. 2014, p.4):

- na região centro-oeste, o jornal Diário de Goiás apresenta a seção Infográfico¹²¹;
- na região sul, o jornal Zero Hora, de Porto Alegre apresenta o blog Livre Acesso¹²²;
- na região nordeste, o jornal O Povo, no Ceará apresenta a seção Infográfico¹²³;
- na região sudeste, o Correio do Brasil, no Rio de Janeiro, apresenta a seção Pesquisa e Infográficos¹²⁴;
- também na região sudeste, em São Paulo, os quatro veículos visitados apresentaram seções dados ou infografia;
- Folha de S. Paulo – blog Folha SP Dados, que se encontra desativado¹²⁵;
- o portal G1 – apresenta a seção Infográficos¹²⁶;
- Estado de S. Paulo – apresenta a seção Infográficos¹²⁷ e o blog Estadão Dados¹²⁸;
- portal IG – apresenta a seção especial Infográficos¹²⁹.

Foram verificadas nesta amostra um total de 8 veículos, sendo 1 da região centro-oeste, 1 da região sul, 1 da região nordeste, 5 da região sudeste sendo que destes, 4 estavam na capital paulista. Isso não quer dizer que não existam reportagens, infografias, oriundas de grandes bases de dados em outras seções. Por isso, a partir

¹²¹ <http://diariodegoias.com.br/infograficos>

¹²² <http://wp.clicrbs.com.br/livreacesso/>

¹²³ <http://www.opovo.com.br/infograficos/>

¹²⁴ <http://correiodobrasil.com.br/noticias/pesquisas-e-infograficos/>

¹²⁵ <http://folhaspdados.blogfolha.uol.com.br> (desativado)

¹²⁶ <http://g1.globo.com/infograficos/index.html>

¹²⁷ <http://www.estadao.com.br/infograficos/>

¹²⁸ <http://blog.estadaodados.com>

¹²⁹ <http://especiais.ig.com.br/infograficos/>

deste cenário nacional, foi buscada a Associação Brasileira de Jornalismo Investigativo (Abraji) que realiza um levantamento de reportagens guiadas por dados publicadas no Brasil. A iniciativa da Abraji visa criar um banco público de referências nacionais. Qualquer pessoa pode contribuir inserindo reportagens em uma planilha aberta¹³⁰, com conteúdos oriundos de diferentes plataformas incluindo sites, revistas, jornais impressos, entre outros. Nas contribuições feitas até o dia 13 de outubro de 2014 constavam 63 reportagens entre os anos de 2009 e fevereiro de 2014. Em relação às editorias especiais voltadas para jornalismo de dados, a Abraji identificou 5 jornais, sendo 3 da região sudeste (O Globo, Estadão Dados e Folha de S. Paulo), um da região sul (Zero Hora) e um da região nordeste (Gazeta do Povo). O que foi perceptível, a partir do estudo destas reportagens, é que diversas eram baseadas em dados, mas não em base de dados.

O que se observa no jornalismo de dados é o uso de bases de dados públicas, privadas, próprias das organizações jornalísticas como centro da criação. Conteúdos que definem a estrutura, a organização e a apresentação dos conteúdos jornalísticos, a partir de categorias específicas. Bases de dados que permitem a atualização e publicação. Como os pesquisadores Suzana Barbosa e Vinícius Torres apontam, um jornalismo “que demarca a ampliação das possibilidades de emprego das bases de dados no processo de produção de conteúdos jornalísticos, no seu consumo e circulação” (BARBOSA & TORRES, 2013, p. 154).

Há diversas ferramentas sendo desenvolvidas para cada vez mais ter precisão no resultado de buscas, principalmente utilizando os recursos da *Web Semântica* (WS). Os caminhos percorridos para se chegar a compra de um determinado produto, quanto tempo se demora em cada site, que tópicos foram acessados, as conversas em redes sociais digitais. Assuntos que podem ser coletados e mapeados tendo as mais diferentes fontes no universo *web* onde toda e qualquer ação no ambiente digital deixa rastros.

Como afirma Berners-Lee (1998), a *web* semântica é um “esforço” que o homem faz para ajudá-la – a máquina – a compreender os documentos na rede e em rede. A WS tem sido nomeada de diversas formas, sendo a definição de “plataforma” a mais comumente encontrada na literatura científica e nos meios de comunicação especializados em tecnologia e

¹³⁰ A planilha está disponível no endereço http://www.abraji.org.br/?id=90&id_noticia=2717

informática, ao lado de “conjunto de aplicativos inteligentes”, “tecnologias inteligentes” e “extensão da *web* atual” (BERTOCCHI, 2013, p. 135).

Na linguagem humana uma palavra pode assumir diversos significados, mas na linguagem dos computadores essa ambiguidade não está prevista. Os *softwares* computacionais não conseguem fazer tantas distinções e para que se consiga são necessários metadados, anotações semânticas. “(...) as camadas de antenarração do sistema narrativo que trabalham com metadados e as linguagens de marcação podem proporcionar um ambiente digital jornalístico “desambiguado”, ou seja, contribuem para o fortalecimento de uma “*Web Semântica*” (BERTOCCHI, 2013, p. 136). O desafio dela é “fornecer uma linguagem que expresse tanto dados quanto regras de raciocínio sobre estes dados e que permita que as regras existentes sejam entendidas por qualquer sistema de representação de conhecimentos presentes na Web”. (BERNERS-LEE, HENDLER & LASSILA, 2001)¹³¹.

Para Tim Berners-Lee, James Hendler e Ora Lassila, autores do artigo *The Semantic Web - A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities* (2001)¹³², a percepção das potencialidades da *web* semântica acontecerá a partir da apropriação e criação de muitos programas, por parte das pessoas, que resultem em coletas de conteúdos da *web* a partir de diversas fontes. Onde ocorra o processamento da informação e sejam feitos intercâmbios dos resultados com outros programas.

Ao falar de jornalismo de dados, jornalismo em base de dados e visualização de dados, ainda abordamos um universo restrito em usos cotidianos no cenário brasileiro. Por mais que a programação esteja presente em todo o processo informacional (apuração, produção e apresentação), muitas vezes o conhecimento em linguagem de programação e estatística, por parte dos jornalistas, acaba sendo um limitador de pautas diante de um grande volume de dados. Veremos ao longo da pesquisa que a questão do tempo e dos recursos econômicos são também grandes complicadores para

¹³¹ “The challenge of the Semantic Web, therefore, is to provide a language that expresses both data and rules for reasoning about the data and that allows rules from any existing knowledge-representation system to be exported onto the Web” (BERNERS-LEE, HENDLER & LASSILA, 2001). Disponível em: http://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf. Acesso em: 20 de julho de 2014.

¹³² Disponível em: http://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf.

o desenrolar em escala da prática. Por enquanto vamos nos ater a um dos aspectos desta tríade (que não necessariamente é o maior problema).

O jornalista tem duas grandes barreiras de linguagem a transpor. Uma é a da estatística e da capacidade de interpretar dados. A outra é a da competência técnica, ou seja, o jornalista precisa aprender a escrever código. Todo jornalista precisa entender, mesmo que em um nível elementar o que é um código, qual a sua função e como se comunicar com gente que entende da coisa (ANDERSON, BELL & SHIRKY, 2012, p. 50).

Há diversos programas que surgiram nos últimos anos para auxiliar no trabalho de coleta dos dados, assim como diversos cursos online e presenciais para capacitar os profissionais envolvidos¹³³. O Instituto Brasileiro de Pesquisa e Análise de Dados divulgou uma lista com ferramentas para análise de redes e grafos em mídias sociais¹³⁴. O *Outwit Hub*¹³⁵, por exemplo, é utilizado para realizar a raspagem de dados e o *Google Refine*¹³⁶ para a limpeza dos dados. No entanto, mesmo com a presença de aplicativos e ferramentas, é preciso um mínimo de conhecimento, da expertise jornalística para encontrar o que rende boas pautas, fazer as perguntas certas aos dados e de entender que mesmo as ferramentas também possuem limitações. O *Outwit Hub* e o *Google Refine*, por exemplo, tornam-se problemáticos quando se precisa fazer uma tarefa grande e repetitiva por semanas ou meses.

Como visto, há diversas maneiras de manusear os dados que vão além do uso de ferramentas e aplicativos. Há hierarquias de processos que podem ser automatizados por meio de programação de robôs. O jornalista Rodrigo Burgarelli, do blog Estadão Dados, defende a existência de pelo menos três tipos de usos de programação no jornalismo: o primeiro é a análise de dados, que é a compreensão do que eles realmente significam e onde muitas vezes é utilizada a linguagem R¹³⁷. O segundo uso diz respeito a raspagem ou *scrapping* de dados, que é a captura de informações não estruturadas que estão em páginas da Internet e a organização das mesmas em banco de dados ou tabelas. Neste momento, tem-se utilizado com frequência a linguagem

¹³³ Abordaremos um pouco mais sobre eles no V capítulo desta pesquisa.

¹³⁴ A quem interessar, vale uma visita à página uma vez que o objetivo desta pesquisa é discutir outras questões que englobam os processos da visualização de dados. Disponível em: http://blog.ibpad.com.br/index.php/2015/12/26/ferramentas-para-analise-de-redes-e-grafos-em-midias-sociais/?utm_medium=redesads&utm_source=fb&utm_campaign=arssp01&utm_content=65. Acesso em 7 de janeiro de 2016.

¹³⁵ <https://www.outwit.com/products/hub/>

¹³⁶ Informações disponíveis no blog Online Journalism, em: <https://onlinejournalismblog.com/tag/google-refine/>.

¹³⁷ R é uma linguagem e também ambiente de desenvolvimento integrado para cálculos estatísticos e gráficos.

Python. Por fim, o terceiro uso que diz respeito a visualização, que é a apresentação deste trabalho ao leitor. Esta apresentação pode ser feita via infográficos, aplicativos interativos, que podem ser feitos com HTML ou Java Script (BURGARELLI apud FREITAS & PINTO, 2015)¹³⁸. Para Burgarelli, ao ter noções sobre programação, a pessoa entende a lógica e o conceito da atividade passando a ter mais ideias de trabalhos.

O programador Luciano Ramalho, do Garoa Hacker Clube¹³⁹, que tem grande proximidade com o jornalismo por ter passado pelas redações do Brasil Online, UOL, AOL Brasil, defende que mesmo os jornalistas que não trabalham diretamente com jornalismo de dados devem atentar para o cenário tecnológico contemporâneo. Devem se preocupar com formatos e armazenamento dos seus textos. Para ele, o jornalista deve escrever pensando que o texto será um dado para diversos formatos, com uma produção e armazenamento de informação compatível para qualquer fim. Apesar de não existir um padrão, a escrita digital em notação é menos perecível do que a feita em plataformas que podem ter sobrevivência datada, como no *World* ou *Facebook*. “A obsolescência programada, por exemplo, pode ocorrer com o *software* proprietário, mas o *software* livre sempre terá outras formas de resgate” (RAMALHO, 2015, anotações). Ramalho acrescenta ainda, que para ele “o jornalismo digital no Brasil só não é melhor porque nós não sabemos fazer” (ibidem, 2015).

Experiência narrativa e usabilidade

No dia 6 de abril de 2016 o Instituto Brasileiro de Geografia e Estatística (IBGE) divulgou¹⁴⁰ parte da Pesquisa Nacional por Amostra de Domicílios (PNAD) com dados sobre o uso de tecnologias da informação pelos brasileiros. De acordo com a pesquisa, o acesso à Internet pelo celular subiu, de 53% em 2013 para 80% em 2014. Já o acesso a Internet pelo computador apresentou queda, em 2013 era 88% e em 2014 diminuiu para

¹³⁸ <http://portalintercom.org.br/anais/nacional2015/resumos/R10-3036-2.pdf>

¹³⁹ <https://g1.globo.com/jornal-hoje/noticia/2016/04/internet-pelo-celular-ultrapassa-acesso-computador-segundo-ibge.html>

¹⁴⁰ Disponível em: <http://g1.globo.com/jornal-hoje/noticia/2016/04/internet-pelo-celular-ultrapassa-acesso-computador-segundo-ibge.html>. Acesso em 5 de maio de 2016.

76%. Pela primeira vez nas pesquisas brasileiras o acesso à Internet pelo telefone celular ultrapassou o acesso via computador¹⁴¹.

O desafio das redações jornalísticas contemporâneas são inúmeros. Não defendemos aqui que jornalistas devam programar códigos ou que as máquinas contarão histórias melhor que os repórteres, mas as possibilidades de contar as histórias foram ampliadas nos meios digitais. Isto significa experimentações e diferentes usos em oportunidades de comunicar melhor. São pelo menos quatro os canais por onde passam as narrativas (seja a criação de produtos ou serviços): primeiramente pelo modelo tradicional impresso que permanece em exercício; em segundo pela *web* nas telas do computador; em terceiro pelos dispositivos móveis como as telas de aparelhos celulares e em quarto pelas telas de dispositivos móveis de maior porte, como os tablets. São diversos os caminhos para manter e atrair um público leitor a partir de formatos ricos que envolvem narrativas com maior interatividade, muitas inseridas em processos de dados e metadados, como é o caso do foco desta pesquisa (abordaremos melhor este assunto no próximo capítulo).

Concordando com as pesquisadoras da Universidade de São Paulo, Daniela Bertocchi, Stefanie Silveira e Isadora Camargo (BERTOCCHI et. al 2015), no jornalismo, a experiência narrativa começa a ser discutida quando são definidas estratégias de produção de formatos para múltiplos dispositivos. São necessárias soluções de usabilidade e design para os produtos levando em consideração não somente os elementos textuais presentes nas interfaces gráficas, mas também a arquitetura de informação que será decisiva na experiência do usuário final. A narrativa passa a ser vista também como uma articulação de dados e metadados em formatos com um propósito determinado. Não é uma junção de camadas computacionais soltas, pelo contrário, possui camadas que são interligadas. A camada dos dados, da database, é uma camada estrutural, mas não é a única a definir as regras do sistema. “A narrativa como sistema é um conjunto complexo, artificial, no qual atuam distintos atores humanos e não-humanos e que produz um todo maior que suas partes” (BERTOCCHI et al. 2015, p. 79). Sendo necessária a compreensão das suas camadas e dos elementos

¹⁴¹ Com relação ao número de casas com acesso à Internet, em 2013, este número correspondia a pouco menos da metade. Já em 2014, quase 55% das casas tinham acesso à Internet.

contidos no processo de contar as histórias, de trabalhar o elo entre o conteúdo, os usuários e a cognição.

Fica ainda o dilema sobre o que fazer diante dos diversos canais: a melhor experiência para o usuário é aquela que oferece uma interface digital adaptada a cada tela, a chamada responsividade ou interfaces digitais distintas e elaboradas para cada um dos dispositivos móveis? Agora, se por um lado temos a questão de mudanças no fazer e pensar jornalístico agregando cada vez mais recursos computacionais, por outro também estão presentes as questões políticas e de regulamentação da Internet e uso de dados no Brasil. Além da usabilidade, regulamentações das conexões em rede e as políticas de transparência interferem no incentivo, no letramento digital e nas apropriações dos dados. Impactos em toda a sociedade. Privacidade, votação do Marco Civil e neutralidade da rede são algumas dessas discussões e conectar estes assuntos nos levam a uma maior compreensão sobre o terreno e as diversas facetas da visualização de dados no jornalismo brasileiro.

Cenário político e discussões sobre a Internet

Quais as principais discussões sobre o uso de dados no Brasil? Como elas impactam no jornalismo de dados e na visualização de dados? O Brasil tem seus diferenciais em relação a outros países que devem ser abordados para a melhor compreensão do todo e que interferem no objeto de estudo proposto nesta pesquisa. Nesta paisagem onde a digitalização modifica a maneira como as pessoas se relacionam com outras pessoas, com produtos e com conteúdos; onde empresas buscam mapear o seu público consumidor; instituições acadêmicas visam conhecer os atores sociais e seus comportamentos a partir do digital, governos em todo o mundo tendem a aumentar a transparência disponibilizando dados referentes à administração pública sob licenças abertas (*open data*). No Brasil, o acesso a informação está previsto na Constituição Federal e na Declaração Universal dos Direitos Humanos. Em 18 de novembro de 2011 foi sancionada a lei federal brasileira número 12.527 conhecida como Lei de Acesso à Informação (LAI)¹⁴². Para o pesquisador Marcos de Jesus Bonfim (2014), da Universidade de São Paulo, ela se configura como um marco jurídico e institucional

¹⁴² <http://www.acessoainformacao.gov.br/assuntos/conheca-seu-direito/a-lei-de-acesso-a-informacao>

que renova princípios da Administração Pública para uma cultura de transparência do setor público com a sociedade civil. Esta lei visa permitir a qualquer cidadão brasileiro consultar órgãos públicos sobre informações e dados de seu interesse. Como complementa a pesquisadora Mariângela Furlan Haswani (2006), também da Universidade de São Paulo, há demandas internas e externas de trabalhos que envolvem grande números de dados.

A demanda existe: no ambiente externo, parte substancial do conteúdo informativo dos meios de comunicação de massa é ocupada com assuntos que envolvem os órgãos públicos estatais. Política, economia, transportes, educação, segurança pública, entre outros, vem sendo temas prediletos da mídia, variando sua ênfase conforme o perfil de cada audiência. No ambiente interno, as informações se perdem nos meandros da burocracia, não circulam entre funcionários que, nos balcões e nos terminais telefônicos, não sabem o que dizer aos cidadãos. (HASWANI, 2006, p. 27).

A partir da Lei de Acesso à Informação foi instituído o Portal Brasileiro de Dados Abertos¹⁴³, uma ferramenta disponibilizada pelo governo para que qualquer pessoa possa encontrar dados e informações públicas. O portal tem o objetivo de divulgar todo e qualquer dado e funciona como um catálogo que facilita a busca e o uso de dados publicados pelos órgãos do governo. O plano estratégico do governo federal prevê que em poucos anos o portal tenha dados dos órgãos federais e das esferas estaduais e municipais¹⁴⁴. Em 2007, especialistas em dados abertos da OpenDataGov¹⁴⁵ estabeleceram oito princípios que definem os dados abertos governamentais¹⁴⁶. São eles:

1. Completos: Todos os dados públicos estão disponíveis. Dado público é o dado que não está sujeito a limitações válidas de privacidade, segurança ou controle de acesso.
2. Primários: Os dados são apresentados tais como coletados na fonte, com o maior nível possível de granularidade e sem agregação ou modificação.
3. Atuais: Os dados são disponibilizados tão rapidamente quanto necessário à preservação do seu valor.

¹⁴³ Disponível em: <http://dados.gov.br>

¹⁴⁴ Informação obtida na página do portal, em <http://dados.gov.br>. Acesso em 7 de maio de 2016.

¹⁴⁵ Disponível em: <https://opengovdata.org/>. Acesso em 25 de novembro de 2015.

¹⁴⁶ Pode-se dizer ainda que dados abertos governamentais são dados que são disponibilizados na *WEB* pelo governo em formato aberto. Entende-se formato aberto como formato que pode ser manipulado e lido por máquina; são considerados formatos abertos: txt; xml; csv. Disponível em: http://www.cps.sp.gov.br/pos-graduacao/workshop-de-pos-graduacao-e-pesquisa/008-workshop-2013/trabalhos/gestao_estrategica_de_tecnologia_da_informacao/121301_622_633_FINAL.pdf. Acesso em: 8 de março de 2015.

4. Acessíveis: Os dados são disponibilizados para o maior alcance possível de usuários e para o maior conjunto possível de finalidades.
5. Compreensíveis por máquinas: Os dados são razoavelmente estruturados de modo a possibilitar processamento automatizado.
6. Não discriminatórios: Os dados são disponíveis para todos, sem exigência de requerimento ou cadastro.
7. Não proprietários: Os dados são disponíveis em formato sobre o qual nenhuma entidade detenha controle exclusivo.
8. Livres de licenças: Os dados não estão sujeitos a nenhuma restrição de direito autoral, patente, propriedade intelectual ou segredo industrial. Restrições sensatas relacionadas à privacidade, segurança e privilégios de acesso são permitidas.

Temos portanto um grande volume de informações, que tende a ser cada vez mais acessível. Este é um ponto da rede, mas como já abordado neste capítulo, toda e qualquer ação no ambiente digital deixa rastros e é passível de ser coletada. O rápido crescimento das redes sociais digitais e os avanços da tecnologia vêm permitindo o processamento em tempo real de uma grande quantidade de dados. São muitos os usos e finalidades de estar inserido no ambiente digital, seja por trabalho, por estudo ou por relações pessoais.

Liberdade e privacidade na rede

Há diversos aspectos a serem abordados ao falar sobre liberdade e privacidade na rede. As discussões estão em andamento, as regulamentações em definição e estas são tendências mundiais. Abordaremos, de forma breve, algumas das temáticas que devem ser levadas em consideração para contextualizar e situar o cenário de dados brasileiros, no qual o jornalismo de dados está inserido. São questões que envolvem usabilidade, acessibilidade a rede e que impactam na presença da visualização de dados no jornalismo.

A vida nas redes da *web* é de exposição, muitas vezes ilimitada em relação a intimidade. O que é realmente privado na *web*? As pessoas estão aprendendo, na prática, o que deve ou não ser mostrado de cada um e como lidam com a exposição ou

melhor, com a super-exposição, com as questões de intimidade e gerenciamentos de crises pessoais em grande escala.

(...) o cyberspace (...) potencializou (...) as possibilidades de inserção de vozes diferentes no contexto público. Pessoas comuns, longe da missão de trabalho e a título de livre expressão do pensamento, providas ou não de interesses inconfessos, põem-se em contexto público, sob o usufruto da democracia como valor universal, aparentemente sem qualquer preocupação com a natureza privada ou pública dessa atuação. (TRIVINHO, 2014, p. 238).

Paralelamente a esta vivência e experiência coletiva que envolve o fórum íntimo e privado, há outras vertentes da questão. Elas dizem respeito ao controle, supervisão e acesso às informações trafegadas na *web*. As justificativas para as ações de controle e supervisão giram em torno da segurança e das reservas de mercado. Em prol de combater os males de determinadas práticas na Internet, entre eles o terrorismo e as redes de pedofilia, são buscadas por diversas lideranças a aprovação de projetos de controle dos conteúdos e da navegação na Internet (AMADEU, 2012, p. 61). Esta supervisão e controle de dados também é contestada por setores econômicos da sociedade, como por exemplo, as indústrias de intermediação afetadas pela expansão e proliferação de conteúdos na *web* (indústrias fonográfica, cinematográfica e editorial, os radiodifusores e as companhias de telecom ligados muitas vezes a manutenção dos sistemas de propriedade intelectual).

Em uma apresentação realizada em agosto de 2014¹⁴⁷, o pesquisador brasileiro e doutor em política Sérgio Amadeu levantou questões importantes sobre a liberdade e a privacidade na Internet. Para ele, a Internet é uma rede cibernética, simultaneamente de comunicação e controle, onde a liberdade de expressão depende da existência da neutralidade de rede, da inexistência de censura prévia e do anonimato na navegação. Para ele, na comunicação em redes digitais, quem controla a infraestrutura pode controlar os fluxos de informação, uma vez que as arquiteturas de rede são arranjos de códigos, definições sobre como proceder para realizar a comunicação entre computadores. Desta forma é possível limitar e/ou facilitar processos de comunicação, assim como bloquear ações e conteúdos. Para Amadeu, a liberdade de expressão na Internet é bem maior do que a liberdade no *Facebook*, por exemplo, que

¹⁴⁷ https://samadeu.files.wordpress.com/2010/03/01_apresentacao_sergioamadeu_egibr.pdf

não está limitada somente pelas definições declaradas em documentos de políticas de uso de dados.

O modo como a Internet foi construída beneficia a diversidade cultural. A arquitetura da rede é distribuída, não existindo centros obrigatórios de tráfego dos fluxos de informação. Ela é reconfigurada constantemente pelos usuários, fruto de uma construção tecnológica aberta. “Até o momento, enquanto prevalecer o princípio da neutralidade da rede, não é necessário se subordinar a determinados formatos ou aceitar conteúdos e doutrinas específicos para participar da Internet” (AMADEU, 2012, p. 63).

A definição de como será a Internet em um futuro próximo dependerá do resultado do embate atual a respeito do princípio da neutralidade da rede. Para compreender bem a dimensão deste contencioso que envolve usuários e pioneiros da Internet, de um lado, contra as operadoras de telecomunicações e a indústria de *copyright*, de outro, é importante observar como a rede mundial de computadores se estrutura e como ela tem funcionado até hoje (AMADEU, 2012, p. 63).

O princípio da neutralidade da Internet se dá pelo seu caráter diante da independência das camadas da rede. A Internet se configura como uma rede de redes. Estas camadas são interligadas, mas completamente independentes umas das outras. As camadas físicas possuem protocolos que permitem que os *bits* sejam transformados em sinais elétricos que são enviados pelas redes físicas controladas pelas operadoras de telecomunicações. Já as camadas lógicas se constituem como de enlace, de rede, de transporte e de aplicação (AMADEU, 2012). Como todo o desenvolvimento da arquitetura da Internet foi pensado para que a inteligência da rede e o gerenciamento humano das aplicações estivessem nas “mãos dos usuários finais” e não sob o controle dos intermediários, o princípio da neutralidade da rede também assegura um outro princípio denominado *end-to-end*, onde “quem deve definir o que deve ser comunicado são as pontas comunicantes da rede e não os meios que estão entre eles”. (AMADEU, 2012, 72).

A independência de uma camada diante da outra assegura o dinamismo da rede. Para se criar uma nova aplicação, não é necessário obter autorização de controladores de quaisquer outras camadas, nem de órgãos públicos, nem de empresas privadas. Por isso, a criatividade é tão intensa na Internet. Qualquer um pode criar novos conteúdos, novos formatos e novas tecnologias. Novos protocolos de comunicação também podem ser criados e obter um enorme sucesso. Para tal eles devem interagir com os demais

protocolos essenciais das outras camadas. Até agora, a neutralidade de uma camada diante da outra tem sido um dos principais fatores que garantem a cultura de liberdade na rede. (AMADEU, 2012, p. 71).

No Brasil a Lei 12.965/14¹⁴⁸, conhecida como o Marco Civil da Internet, foi sancionada pela presidente Dilma Rousseff no dia 23 de abril de 2014, entrando em vigor em junho do mesmo ano. A lei regulamenta o uso da Internet no país por meio da previsão de princípios, garantias, direitos e deveres do uso da rede, assim como as diretrizes para a atuação do Estado. O texto estabelece que o Comitê Gestor da Internet (CGI)¹⁴⁹ é o responsável por determinar as diretrizes. As regras que regulamentam a lei foram elaboradas por meio de quatro consultas públicas elaboradas pela Agência Nacional de Telecomunicações, CGI e Ministério da Justiça. A ideia do projeto de lei do Marco Civil surgiu em 2007 em função da resistência social ao projeto de lei de cibercrimes conhecido como Lei Azeredo, muito criticado por ser um AI-5¹⁵⁰ digital.

O “Marco Civil da Internet” é um projeto de lei que visa a consolidar direitos, deveres e princípios para a utilização e o desenvolvimento da Internet no Brasil. A iniciativa partiu da percepção de que o processo de expansão do uso da Internet por empresas, governos, organizações da sociedade civil e por um crescente número de pessoas colocou novas questões e desafios relativos à proteção dos direitos civis e políticos dos cidadãos. Nesse contexto, era crucial o estabelecimento de condições mínimas e essenciais não só para que o futuro da Internet seguisse baseado em seu uso livre e aberto, mas que permitissem também a inovação contínua, o desenvolvimento econômico e político e a emergência de uma sociedade culturalmente vibrante (CGI, 2015)¹⁵¹

No artigo três¹⁵² a lei garante a liberdade de expressão, comunicação e manifestações do pensamento, proteção dos dados pessoais, na forma da lei, preservação e garantia da neutralidade de rede. O artigo 10 define que a guarda e a disponibilização dos

¹⁴⁸ Disponível em: http://www.planalto.gov.br/CCIVIL_03/_Ato2011-2014/2014/Lei/L12965.htm. Acesso em 22 de janeiro de 2014.

¹⁴⁹ <http://www.cgi.br/>

¹⁵⁰ Ato institucional do regime militar que esteve em vigor de 1968 a 1978 e que dava poderes extraordinários ao presidente da república e suspendia várias garantias constitucionais.

¹⁵¹ Disponível em: <http://www.cgi.br/publicacao/o-cgi-br-e-o-marco-civil-da-internet/>.

¹⁵² De acordo com o artigo 3º, da lei Marco Civil brasileiro, a disciplina do uso da Internet no Brasil tem os seguintes princípios:

- I - garantia da liberdade de expressão, comunicação e manifestação de pensamento, nos termos da Constituição Federal;
- II - proteção da privacidade;
- III - proteção dos dados pessoais, na forma da lei;
- IV - preservação e garantia da neutralidade de rede;
- V - preservação da estabilidade, segurança e funcionalidade da rede, por meio de medidas técnicas compatíveis com os padrões internacionais e pelo estímulo ao uso de boas práticas;
- VI - responsabilização dos agentes de acordo com suas atividades, nos termos da lei;
- VII - preservação da natureza participativa da rede;
- VIII - liberdade dos modelos de negócios promovidos na Internet, desde que não conflitem com os demais princípios estabelecidos nesta Lei.

registros de conexão, de acesso, assim como dados pessoais devem atender à preservação da intimidade. No artigo 7, a lei determina que o provedor responsável pela guarda somente disponibilizará os registros que possam contribuir para a identificação do usuário ou do terminal, mediante ordem judicial.

Neste período de definições e usabilidades desta nova ambiência, há diferentes interpretações de situações corridas no ambiente digital e inúmeras contestações diante das regulamentações que aparecem no cenário nacional e internacional. Por exemplo, empresas como *Apple*, *Google* e *Facebook*, assim como organizações não-governamentais solicitaram ao Congresso dos Estados Unidos regras mais claras e maior transparência sobre pedidos de informação em que as autoridades invoquem razões de segurança. Estas empresas procuraram tranquilizar os usuários após as denúncias feitas pelo analista informático Edward Snowden de que os serviços de segurança e espionagem dos Estados Unidos tem acesso a informações “particulares”. No Brasil, até maio de 2016, por duas vezes o aplicativo de mensagem instantânea *WhatsApp* foi bloqueado por determinação judicial. No dia 2 de maio de 2016, por exemplo, a justiça do Estado de Sergipe determinou que as operadoras de telefonia do país bloqueassem o acesso de todos os brasileiros ao aplicativo por um prazo de 72 horas porque o *Facebook*, dono do *WhatsApp*, não cumpriu uma decisão judicial anterior de compartilhar informações que subsidiariam uma investigação criminal. No dia seguinte, dia 3 de maio, o *WhatsApp* conseguiu uma decisão favorável da Justiça e o serviço passou a ser restabelecido assim que as operadoras foram notificadas e fizeram ajustes em sua rede de telefonia.

O cenário é complexo, está em andamento e contínua transformação. Há discussões sobre o limite de dados da Internet, por exemplo, onde as operadoras que atuam no Brasil querem oferecer planos de internet fixa (usada nas residências e empresas) com limite de download. Onde o serviço pode ser suspenso quando o usuário atingir uma determinada quantidade de arquivos e dados baixados¹⁵³. As discussões acontecem em diversos âmbitos, seja o acadêmico, o jurídico, o político, entre outros. Fato é que a vivência digital está inserida no cotidiano, governos em todo o mundo tendem a aumentar a transparência disponibilizando cada vez mais dados referentes à

¹⁵³ A Mentira Por Trás do Limite da Internet – Marcelo Toledo - <https://www.facebook.com/MarceloToledoDONO/videos/1209510815733649/?pnref=story>.

administração pública. A produção de conteúdos jornalísticos e seções em sites especificamente para disponibilizar dados produzidos, organizados e estruturados por jornalistas através de APIs é uma realidade. Cenário que para nós confirma a expansão e relevância do jornalismo de dados e das visualizações de dados no jornalismo digital. Como fazer jornalismo de dados, com visualizações de dados, arquivos leves, responsivos diante de um cenário de acesso limitado de banda larga, acessos crescentes a Internet por aparelhos móveis, tendências de feeds¹⁵⁴ de notícia são, entre inúmeras outras variáveis, desafios que permanecem para o jornalismo contemporâneo, sem fórmulas mágicas ou consolidadas de sucessos reproduzíveis.

Considerações

O mapa da cólera em Londres, desenvolvido pelo médico Jonh Snow, em 1854, foi um marco de como uma infografia baseada em dados contribuiu para a compreensão de um fenômeno da época. A partir da ciência por trás dos dados e da forma de apresentação da informação foi possível compreender o meio transmissor da doença que tinha levado tantas pessoas à morte. Nos dias atuais a visualização de dados continua sendo um importante meio de investigação e compreensão de fenômenos. No entanto, hoje falamos de um volume, variedade e velocidade de dados humanamente imensuráveis. A sociedade está cada vez mais digitalizada, dados em diferentes extensões são passíveis de serem armazenados e correlacionados. Neste cenário falamos de um jornalismo de dados. Um jornalismo que visa o entendimento de temas complexos de relevância social e que sintoniza o jornalismo com as novas necessidades de compreensão informativa da sociedade contemporânea.

O segundo capítulo teve por objetivo exemplificar o jornalismo de dados e a visualização para, a partir da sua presença no cotidiano discutir com mais profundidade as questões teóricas que serão trabalhadas no capítulo III. Os exemplos envolveram investigações e inúmeras correlações de dados que foram apresentadas aos leitores de maneiras distintas como em infográficos e mapas, tendo como foco o

¹⁵⁴ “É um formato de dados usado em formas de comunicação com conteúdo atualizado frequentemente, como sites de notícias ou blogs. Distribuidores de informação, blogueiros ou canais de notícias disponibilizam um feed ao qual usuários podem se inscrever, no formato de um link. Outros formatos de dado possíveis de serem comunicados por *feeds* são arquivos de áudio, podcasts e vídeos”. Disponível em: <https://pt.wikipedia.org/wiki/Feed>. Acesso em 5 de maio de 2016.

cenário brasileiro. Distinguimos em três principais áreas como as visualizações de dados vem aparecendo no país: nos veículos tradicionais de mídia, nas iniciativas independentes e nos laboratórios acadêmicos. Por isso, trouxemos exemplos que foram coletados ao longo dos primeiros 15 meses de observação da prática em âmbito nacional. De imediato percebemos que a formação das equipes, as fontes dos dados e o tempo de produção dos conteúdos são distintos de outras práticas jornalísticas, frutos de um sistema narrativo que na contemporaneidade adquire novas rotinas produtivas. As equipes são formadas por profissionais de competências distintas, principalmente jornalistas, programadores, designers, estatísticos. As fontes envolvem bases de dados, muitas originárias de serviços públicos, mas também não são descartadas entrevistas e outros recursos já tradicionais à prática jornalística como meio de construção das peças. Por conta do aprofundamento dos temas, do manuseio com os dados e das apresentações dos conteúdos (seja para a interpretação dos próprios repórteres ou para o público/leitor), são trabalhos que podem demorar de poucas horas há meses.

Em um segundo momento a preocupação foi pontuar as discussões políticas e leis de transparência que interferem na produção e disseminação da usabilidade na *web* e consequentemente da prática do jornalismo de dados no país. O contexto nacional visa o estabelecimento de condições mínimas para um uso da Internet de forma livre e aberta, permitindo inovações, desenvolvimento econômico e político, mas não deixa de ser um cenário que está em andamento e contínua transformação. Há tensões sobre o controle de tráfego de dados, privacidade na rede, fatores constantes de discussão, consultas públicas, que envolvem diretamente o estudo desta pesquisa. Ter acesso limitado a download, por exemplo, pode fazer com que leitores optem por não baixar conteúdos jornalísticos mais pesados, não consultar bases de dados, implicações que podem vir a interferir em um modo de fazer jornalismo que possui relevância social, mas que também precisa ser consumido, ser apropriado pelos leitores. Quais caminhos então deverão ser tomados? Há meios de oferecer serviços mais compactos? Que posicionamentos devem ser tomados a partir das decisões políticas e econômicas nacionais que envolvem o uso de dados, as fontes dos dados? Estar atento ao cenário diz respeito ao jornalismo de dados. No próximo capítulo o desafio é situar a visualização de dados dentro do sistema narrativo contemporâneo.

III. O FAZER JORNALISMO DE DADOS E A VISUALIZAÇÃO NO SISTEMA NARRATIVO

Resumo: Neste capítulo nos aprofundaremos nas visualizações em base de dados apontando em que momento esta prática se encontra dentro da produção jornalística. Para tanto, temos a tarefa de abordar conceitos de narrativa como sistema, formato e design no jornalismo. Por fim, apontaremos as questões de credibilidade e valor jornalístico diante do cenário cultural contemporâneo.

Palavras-chave: sistema narrativo, formato, visualização de dados, designer, valor jornalístico

“O futuro já chegou, só não está uniformemente distribuído”. (William Ford Gibson, 1999)¹⁵⁵

Diferentes olhares e novas práticas

Em 2012, os pesquisadores Chris Anderson, Emily Bell e Clay Shirky lançaram o relatório *Jornalismo Pós-Industrial - adaptação aos novos tempos*, pelo *Tow Center for Digital Journalism da Columbia Journalism School*¹⁵⁶. O trabalho abordou o atual estágio do jornalismo, as condições técnicas, materiais e os métodos empregados na apuração e divulgação das notícias até o fim do século 20 e que já não se aplicam nos dias atuais. “Estamos em meio a uma revolução, e a adaptação às novas fronteiras da profissão é a condição de sobrevivência nesse cenário, que prevê o uso intensivo de base de dados, além da interação com múltiplas fontes e com o público” (ANDERSON et al., 2012, p. 30). O relatório partiu de cinco grandes convicções: a) o jornalismo é essencial; b) o bom jornalismo sempre foi subsidiado; c) a Internet acaba com o subsídio da publicidade; d) a reestruturação se faz, portanto, obrigatória; e) há

¹⁵⁵ William Gibson cunhou o termo cyberpunk e é autor da obra *Neuromancer* (1984). Citou a frase no dia 30 de novembro de 1999, em entrevista a NPR Interview.

¹⁵⁶ A revista ESPM publicou o artigo *Jornalismo pós-industrial – adaptação aos novos tempos*. São 60 páginas (sem gravuras) divididas em três partes com o estudo preparado pela Columbia University sobre como a imprensa pode prosperar para além do mercado em crise. Disponível em: http://www.espm.br/download/2012_revista_jornalismo/Revista_de_Jornalismo_ESPM_5/files/assets/common/downloads/REVI STA_5.pdf.

muitas oportunidades de fazer um bom trabalho de novas maneiras. Neste capítulo, em um primeiro momento, vamos nos ater a duas dessas premissas: o jornalismo é essencial e há muitas oportunidades de fazer um bom trabalho de novas maneiras. A partir do que o jornalismo significa enquanto essência é possível discutir as suas técnicas. Aqui optamos por trabalhar com o lado sério do jornalismo, o que pode ser chamado por *hard news* ou *accountability journalism*¹⁵⁷; coberturas de fatos importantes e reais que impactam e podem mudar rumos sociais.

O jornalismo expõe a corrupção, chama a atenção para a injustiça, cobra políticos e empresas por promessas e obrigações assumidas. Informa cidadãos e consumidores, ajuda a organizar a opinião pública, explica temas complexos e esclarece divergências fundamentais. O jornalismo exerce um papel insubstituível tanto em regimes democráticos como em economias de mercado. (ANDERSON et al., 2012, p. 33).

Para nós o jornalista não é somente um narrador de fatos, ele é o responsável por coletar, apurar, interpretar e contextualizar fatos. Diante do cenário tecnológico pode recorrer a gráficos interativos, solicitar imagens e informação ao público/leitor, disponibilizar acesso direto a bancos de dados e distribuir matéria por redes sociais. “O aspecto mais animador e transformador do atual cenário jornalístico é poder explorar novas formas de colaboração, novas ferramentas de análise e fontes de dados e novas maneiras de comunicar o que é de interesse do público” (ANDERSON et al., 2012, p. 38). O jornalista não foi substituído e nem será extinto. O profissional foi deslocado para um ponto acima na cadeia editorial exercendo a função de verificar, interpretar e dar sentido à enxurrada de texto, áudio, fotos e vídeos inseridos na *web*, tornou-se um curador, alguém capaz de fazer a filtragem, a reorganização e a interpretação do que se procura e é encontrado na Internet (BEIGUELMAN, 2016)¹⁵⁸. Num ambiente em que a informação é base para a produção do saber e conta com a presença de algoritmos para a verificação de padrões coletados a partir de metadados, que registram, por exemplo, os rastros dos indivíduos na *web*. Os algoritmos, diante do volume de dados com que lidamos, permitem que operemos a dadosfera. (BEIGUELMAN, 2016)¹⁵⁹. No entanto, o que ainda se observa na prática cotidiana, é

¹⁵⁷ Traduzindo ao pé da letra: notícias duras e jornalismo de prestação de contas, respectivamente.

¹⁵⁸ O termo foi amplamente discutido no ebook Curadoria Digital e o campo da Comunicação. Disponível para download em <http://grupo-ecausp.com/novo-ebook-curadoria-digital-e-o-campo-da-comunicacao/>

¹⁵⁹ Disponível em: <http://noticias.uol.com.br/opiniaao/coluna/2016/02/21/curadoria-de-conteudo-e-o-lugar-do-humano-na-internet.htm>. Acesso em 21 de fevereiro de 2016.

que diversos profissionais e instituições jornalísticas não tem tirado proveito da explosão de conteúdo potencialmente relevante da comunicação digital.

Devia haver mais trabalho de reportagem, e não menos, e essa reportagem devia aprender a conviver com formas mais recentes de apuração de informações de interesse jornalístico. Reconhecemos que o colapso econômico de jornais representa uma ameaça bastante real para o trabalho de reportagem; a solução desse dilema exigirá uma nova atenção a instituições jornalísticas (...).(ANDERSON et al., 2012, p. 43).

Estamos em processos de mudança, adaptação, inovações que possuem o uso intensivo de base de dados, interação com múltiplas fontes e com o público/leitor/usuário. Transformar dados em informação e a informação em conhecimento depende de expertises, de análises, interpretações. Novas possibilidades para o jornalismo exigem novas formas de organização e somente incorporar técnicas não serão suficientes para a adaptação ao novo ecossistema. Em 2011 a professora Elizabeth Saad já fazia o alerta para as empresas sobre a necessidade de “reaprender a construir e a disponibilizar o contexto, a sair da fragmentação noticiosa e rumar para uma condução da audiência na busca de mais e mais informações correlacionadas”. (SAAD, 2011, p. 61). Isso em um cenário onde os exploradores de dados, criadores de visualização de informação possuem um conjunto maior de ferramentas. É preciso que os jornalistas, nas suas mais diferentes funções (repórteres, editores, diretores) tenham um olhar para o *backend*¹⁶⁰ e para o *frontend*¹⁶¹ jornalísticos.

Concordamos com a pesquisadora Daniela Bertocchi que as bases de dados são essenciais, mas não estão no centro do sistema narrativo, uma vez que devem ser considerados os “algoritmos, os sistemas publicadores, os metadados e todas as demais camadas que se apresentam no momento de antenarração são fundamentalmente centrais à formatação da narrativa” (BERTOCCHI, 2013, p. 211). Desta forma, a formatação é o fruto da base de dados e também o resultado do sistema “capaz de revelar comportamento de todo um sistema” (ibidem, 2013, p. 211). Vivenciamos um processo de inovação e absorção de tecnologias nunca experimentados, conduzidos por pessoas de fora da área, com outras expertises; que transmitem informação relevante. Onde se encontra a visualização de dados neste

¹⁶⁰ Interface digital administrativa acessível por administradores do sistema (BERTOCCHI, 2014, p.18);

¹⁶¹ Interface digital acessível por usuários finais (BERTOCCHI, 2014, p.18)

cenário e no sistema narrativo? Como apresentar os dados para o público de maneira inovadora?

Para a pesquisadora Martha Gabriel (2012, anotações)¹⁶² especialista em marketing digital, inventar é criar, engendrar, descobrir. Já inovar é tornar novo, renovar, introduzir novidade em alguma coisa. A invenção tende a ser uma ruptura. A inovação tem como virtude adaptar, substituir, combinar, ampliar ou reduzir, enfim, são outras utilizações. Para a pesquisadora Elizabeth Saad (2014)¹⁶³ inovamos por motivos macros e micros. Macros diante do desperdício de recursos que poderiam ser aproveitados em novos usos, em melhorias em produtos que poderiam ser aplicados em outras funções ou utilidades, em necessidades culturais latentes ou emergentes. Inovamos por motivos micros por novos desenvolvimentos como variável competitiva, liderança, lucratividade, sobrevivência ou reconfiguração de uma marca. A inovação não é uma invenção; ela tem a ver com a criatividade, com soluções, melhorias, criação de novos produtos e usos. No cenário contemporâneo as inovações acontecem nas tecnologias, nas mídias, no jornalismo, nos usuários, na sociedade. Para acontecer depende da cultura e economia da indústria, do processo de gestão da inovação e de pessoas.

O uso de bases de dados reconfigura, pelo viés do cruzamento de dados, um conteúdo mais original ao jornalismo, onde os dados podem ser somados, comparados, integrados entre si. O jornalista, por sua vez, reordena, filtra, segmenta identifica padrões na informação, revela tendências, expõe resultados. Pode participar tanto como produtor do conteúdo como participar do desenvolvimento de *softwares*. A prática do jornalismo digital se dá a partir do agenciamento coletivo de diferentes atores como jornalistas, designers, engenheiros, programadores, arquitetos de informação, especialistas em negócios, usuários, algoritmos, robôs, *softwares*, entre outros. No ecossistema comunicacional, o jornalista assume o processo de humanização dos dados servindo como um investigador, alguém que traduz e conta fatos. Precisa pensar sobre o tipo de experiência narrativa que os usuários poderão atingir ao entrar em contato com as histórias no mundo digital, por diferentes

¹⁶² Módulo de Marketing Digital realizado no primeiro semestre de 2012, no Curso de Gestão Integrada da Comunicação Digital nas Empresas (Digicorp), na Escola de Comunicações e Artes da Universidade de São Paulo.

¹⁶³ Módulo de Tecnologia e Inovação, do Curso de Gestão Integrada da Comunicação Digital nas Empresas (Digicorp), na Escola de Comunicações e Artes da Universidade de São Paulo, 2014. Participação como ouvinte.

dispositivos, e que tipo de reação o sistema narrativo tende a provocar seja em qual tela estiver. “Os formatos mais ricos são aqueles que compreendem de antemão a narrativa como um processo constante de dados, metadados e formatos múltiplos” (BERTOCCHI, 2013, p. 215). E como aponta a pesquisadora Daniela Bertocchi, é exatamente no *backend* jornalístico que ocorrem os agenciamentos das camadas do sistema narrativo. Momento este que possivelmente os usuários finais do produto não terão acesso.

Princípios da narrativa contemporânea

Ao mencionar a importância de narrativas originais e criativas que unem recursos como áudio, vídeo e texto não-lineares, fala-se em hipermídia. Na visão do professor Sérgio Bairon, da Escola de Comunicações e Artes da Universidade de São Paulo, a hipermídia pode ser entendida como “a expressão não linear da linguagem, que atua de forma multimidiática e tem sua origem conceitual no jogo” (BAIRON, 2011, p 7). É uma expressividade da linguagem e não o resultado de uma evolução tecnológica. “Em nível técnico, sua atuação multimidiática significa que, no mesmo ambiente, temos imagens (fixas ou em movimento), textos e sons que sustentam o conteúdo exposto” (ibidem, 2011, p. 8). A hipermídia é um conceito teórico e filosoficamente complexo, onde é preciso ter em mente que as potencialidades são tecnológicas, mas as condições de criação e produção são conceituais. Desafia os paradigmas da Comunicação e da compreensão tradicional no interior da produção de conhecimento e, neste sentido, torna-se fundamental entender o que está além do advento das novas tecnologias; de uma hipermídia não apenas como um aparato, mas sim como um novo meio promotor de conhecimento no qual o jornalismo digital se apropria. Um jornalismo que utiliza narrativas inovadoras, base de dados, imersão, geolocalizadores entre muitos outros recursos disponíveis.

A comunicação em rede acelerou todo o processo de transmissão de informação. Quando transformou o texto impresso em código eletrônico mudou a forma de manipulação e o compartilhamento de dados, possibilitando a construção e habitação de um mundo que proporciona tanto a demonstração de contextos cotidianos quanto a experiência associada a um conceito.

Cada nova máquina tecnosocial contribui com novos dimensionamentos espaciais e temporais que apontam para a multiplicação contemporânea dos espaços, seja como razão primordial do nosso nomadismo estilístico, seja como, multiplicação da necessidade de buscas (BAIRON, 2011, p. 82).

A imersão é uma condição de habitação, revelada pela experiência da construção da relação de sentido entre ambientes e usuários, onde o mais importante na forma de observar o cotidiano não é a descrição dos produtos culturais, mas sim as diversas maneiras que tem os usuários de utilização desses produtos. “(...) dentro do mundo das experiências estéticas, imagens, sons e textos passaram a compor e impor-se, no mundo hipermídia, como elementos conjuntos e cooperativos da produção de sentido e significância reticular”. (BAIRON, 2011, p. 87). Desde a invenção da prensa, a tecnologia da informação ocupou-se de criar e propagar informações em um formato estático, o que permitia que o conhecimento transmitido antes oralmente pudesse agora ser repassado a diferentes leitores em locais e tempos distintos (BARIANI & BAIRON, 2012). No século XIX, a prensa de alta velocidade teve um papel democratizante ao criar e distribuir múltiplas cópias de texto ao invés de restringi-lo, como acontecia na época do manuscrito. Em contrapartida, de acordo o pesquisador George Landow (2009), autor do livro *Hipertexto 3.0*, tal organização hierárquica e linear, típica do texto impresso nem sempre coincide com as necessidades individuais de cada leitor. Para tanto foram inventados os capítulos, números de páginas etc., como forma de melhor direcionar o indivíduo de acordo com seus próprios interesses. A última e grande mudança da tecnologia da informação foi a tecnologia digital.

A hipermídia vai de encontro à dicotomia entre arte, religião e cotidiano e a interpretação. Dicotomias iniciadas no Iluminismo (século XVIII) que definiram a linguagem escrita como a de maior credibilidade em detrimento de outras linguagens e formas de apreensão de conhecimento. Unindo diferentes extensões, a hipermídia traz exatamente na intersecção do objetivo e do subjetivo a busca da verdade e do conhecimento.

(...) o primeiro grande poder definidor da hipermídia está na hibridização das matrizes de linguagem e pensamento, nos processos sígnicos, códigos e mídias que ela aciona e, conseqüentemente, na mistura de sentidos receptores, na sensorialidade global, sinestesia reverberante que ela é capaz de produzir, na medida mesma em que o receptor ou leitor imersivo interage com ela, cooperando na sua realização. Por isso mesmo, em uma

definição sucinta e precisa, hipermídia significa “a interação sem suturas de dados, textos, imagens de todas as espécies e sons dentro de um único ambiente de informação digital”. (FELDMAN, 1995, p. 4).

Para o teórico russo Lev Manovich (2002), o caminho de estrutura de uma narrativa imbrica a experiência do homem com as tecnologias disponíveis. No entanto, ele defende que aliado aos dados computadorizados, a síntese numérica em que as imagens são transformadas devem ser desenvolvidas com estética e elementos poéticos no banco de dados. Manovich aprofunda esta ideia no projeto *Soft Cinemas*¹⁶⁴, onde une trabalhos artísticos de composição das cenas, influências da cultura popular, do filme e da arte. Neste projeto o banco de dados está disponível em diversos locais, como em CD-ROM, DVD, fotografias, possibilitando inúmeras reconstruções narrativas executadas pelo *software* do computador em edição instantânea. Os dados podem ser acessados cronologicamente, por país, artista, possibilitando uma navegação por diferentes caminhos. A tecnologia digital possibilita explorar diversos formatos em narrativas que interagem com o usuário. Não necessariamente a “história” está ali contada, nos casos da hipermídia podem não ter início ou fim, ter o desenvolvimento temático organizado ou não em sequência, “a experiência do usuário com coleções computadorizadas é um pouco distinta da leitura de uma narrativa ou de assistir um filme ou navegar na arquitetura de um site” (MANOVICH, 2002, p. 219). Concordando com a pesquisadora Daniela Ramos (2011), selecionar e combinar bases de dados em formatos digitais, de modo a gerar narrativas é um desafio.

Entendemos, como Manovich (1999, 2006) e Jenkins (2008), que novas formas narrativas não são como as formas clássicas, com trama, enredo e fábula, começo, meio e fim, conectando personagens e situações. Como coloca o primeiro autor, escolher um produto cultural na base de dados do consumo é, de certa forma, contar algo sobre nós mesmos. Para que consigamos contar algo sobre qualquer coisa, precisamos de linguagens e estas, hoje, são selecionadas a partir das bases de dados (RAMOS, 2011, p. 06).

Neste sentido, das linguagens selecionadas por dados, Manovich entende que as características da forma cultural base de dados são (MANOVICH, 2006, p. 63-103):

¹⁶⁴ Disponível em: <http://www.softcinema.net/index.htm?reload>.

1o) Representação numérica – a representação numérica tem em sua essência números, convertidos em código binário, onde são manipulados e capazes de gerar conteúdos com sentido;

2o) Automação – os dados são passíveis de serem programáveis e terem modificação de forma automática, via algoritmos;

3o) Modularidade – os dados são modulares uma vez que a mídia mantém a mesma estrutura em diferentes escalas;

4o) Variabilidade – passível de ser disponibilizado em distintas versões;

5o) Transcodificável – conteúdo passível de ser traduzido para outro formato, de modo programável, sendo codificado através de outros códigos que não são o da sua origem.

Ampliando as perspectivas teóricas, como forma de contextualização, o autor Henry Jenkins (2009) defende que cada vez mais as narrativas estão se tornando construções de universos não explorados ou esgotados em uma única obra ou mídia, o que ele denomina como narrativa transmídia. Em suma, a convergência tecnológica possibilita contar histórias de maneiras inovadoras, unindo modalidades de percepções reais e virtuais tendo a participação do público. Usos de tecnologias que exploram sensações, sentidos. E neste cenário também estão presentes as narrativas jornalísticas diferenciadas, os projetos jornalísticos que envolvem apresentação de uma enorme quantidade de dados. Prática esta que faz parte do jornalismo de dados e que neste artigo aprofundará um formato, o de visualização de dados.

Jornalismo como sistema: seleção de informações e formato narrativo

O pesquisador Ronaldo Henn (2002), retomando os pensamentos do pesquisador Edvaldo Pereira Lima (1982)¹⁶⁵, defende o jornalismo como um sistema construído de

¹⁶⁵ Em dissertação de mestrado defendida na Escola de Comunicação e Artes, da Universidade de São Paulo, em 1982, sob o título “O jornalismo impresso e a teoria geral dos sistemas: um modelo didático de abordagem”.

partes interdependentes, o qual apresenta características definidoras (a linguagem) e desempenha uma função específica (comunicar relatos de acontecimentos sociais). No ambiente digital, a narrativa deixa de ser compreendida como texto e imagens estaticamente diagramados; deixa de ser apreendida como um produto jornalístico final e passa a ser “como um processo dinâmico em constante manifestação e transformação no tempo e espaço digitais”. (BERTOCCHI, 2013, p. 39). O usuário experimenta o fluxo contínuo de eventos.

De acordo com Daniela Bertocchi “a narrativa como sistema é um conjunto complexo, artificial, no qual atuam distintos atores humanos e não-humanos e que produz um todo maior que suas partes” (BERTOCCHI, 2013, p.50). Ela começa a ser construída em cima da codificação maquínica e, conforme as camadas de bases de dados vão dialogando com outras camadas, forma-se um tecido narrativo. Quando a narrativa chega a uma interface para o usuário, ela já foi formatada, renderizada e o seu fluxo continua, uma vez que os usuários acrescentam comentários, compartilham o texto, links. “A narrativa assim manifesta-se a partir da interação entre camadas computacionais, – a strata – e revela-se portanto um processo artificial de design informático” (Ibidem, 2013, p. 55) . A camada mais próxima da superfície é o formato. “Os formatos modelizam as linguagens digitais. Formatos são interfaces para as bases de dados disponíveis no ciberespaço” (RAMOS, 2011, p. 55).

Daniela Bertocchi propôs em sua tese doutoral um modelo de sistema narrativo, explicando em atos o que cada fase representa (BERTOCCHI, 2013, p.70):

- Ato I – Antenarração dos dados – o levantamento e a seleção das informações de interesse público guardadas nas bases de dados;
- Ato II – Antenarração dos metadados – a semantificação destes dados para compreensão de *softwares* e algoritmos, para a apresentação de narrativas nas interfaces digitais e também para a organização interna de dados em sistemas publicadores;
- Ato III – Narração ou formatação narrativa – a organização corporificada da narrativa na interface digital para acesso dos usuários finais.

Figura III. 11 – O Sistema Narrativo no Jornalismo Digital (BERTOCCHI, 2013)

Fonte: BERTOCCHI, 2013, p. 70.

A programação de máquinas e a automatização de processos possibilitam garimpar com rapidez grandes volumes de dados. Algumas relacionadas a *web* semântica (geração de conhecimento pela máquina) podem fazer pequenas análises ou até escrever textos de resultados de jogos apenas atualizando os dados, como é o caso divulgado em 2011, pela empresa *Narrative Science*¹⁶⁶, uma *startup* de Evanston, Illinois (EUA), que apresentou computadores imitando o raciocínio humano ao escrever notícias. O *software* toma dados como os de estatísticas esportivas, relatórios financeiros de empresas, lances e vendas de moradias e os transforma em notícias. No entanto, em todas as outras situações, inclusive para ensinar a máquina, é preciso fazer a leitura dos dados e interpretá-los. Em suma, disponibilidade e acessibilidade são questões diferentes. Neste contexto, é fundamental que os jornalistas compreendam a natureza do que os conjuntos imensos de dados representam e podem

¹⁶⁶ <https://www.narrativescience.com/>

revelar, assim como a importância de ter uma escrita compatível com diferentes formatos.

É na camada de antenarração que ocorrem possivelmente as influências mais significativas da inteligência artificial no sistema narrativo e que, uma vez agenciadas, tendem também a modificar, alterar o sistema narrativo e seu comportamento. A antenarração abre vias para se pensar numa computação social centrada no humano. (BERTOCCHI, 2013, p. 213).

Fazer com que dados de fontes distintas e diferentes extensões dialoguem é um desafio, onde a aproximação da programação com os jornalistas, principalmente na fase de antenarração dos dados é determinante para o desenvolvimento de todo o sistema narrativo. Encontrar variáveis e correlações são uma segunda etapa de trabalho com os dados. Para o avanço neste setor são necessários investimentos em tecnologia, em *softwares* e em treinamento para que os jornalistas aprendam a lidar com os dados. Jornalistas que estejam aptos para a colaboração e para o desenvolvimento de programas sugerindo novas buscas de conteúdos e interações que podem gerar resultados mais interessantes para os usuários finais. O jornalista deve ser capaz de analisar dados e indicadores e ter consciência de que os números representam atividades humanas (ANDERSON et al., 2012). Quanto mais complexos os projetos, maior a combinação de diferentes conhecimentos e habilidades, além de um sistema de produção planejado e hierarquizado em funções de trabalho, “o que dificilmente poderia ser realizado por apenas um profissional utilizando uma ferramenta” (CAIRO, 2008, p. 84).

É cada vez mais comum, por exemplo, a descrição de acontecimentos e compartilhamento feitos por cidadãos conectados à *web*, principalmente em situações de desastres naturais, acidentes, roubos. Neste sentido, a apuração dos fatos ocupa um lugar relevante para o trabalho jornalístico de interesse público uma vez que jornalista não é um mero narrador de fatos. A antenarração de metadados é de origem cognitiva e de geração de sistemas culturais simbólicos. A consolidação da prática da antenarrativa de metadados contribuirá para o desenvolvimento da *web* semântica e da esfera semântica (inteligência coletiva humana); possibilitando pensar em uma computação social centrada no humano. Onde ferramentas, produtos e inclusive construções narrativas levem em conta a experiência do usuário, seus gostos e sua linha de raciocínio. O objetivo será engajar as pessoas pelo que elas gostam e que

possam vir a compartilhar, comentar. São inúmeras as possibilidades no meio digital, mas todas as práticas, sejam de jornalistas, publicitários ou empresas de mídia “devem engajar o público, agregar valor e gerar conhecimento. Devem sempre trazer algo a mais”, disse a professora Dra. Elizabeth Saad, durante o Festival Path 2015, durante a sua fala sobre as tendências na comunicação, na qual ressaltou a *Digital Humanities*. Na sociedade digitalizada, as atividades de pesquisa das áreas das Ciências Humanas e Sociais estão cada vez mais associadas as ferramentas oferecidas pela Computação. As metodologias da História, Filosofia, Artes, estudos culturais, por exemplo aparecem associadas a visualização de dados, recuperação de informação, mineração de dados e estatística. O olhar voltado para os dados aproxima o jornalismo da ciência de dados.

Temos um também complexo *backend* jornalístico que acaba por ser organizado em formatos, uma vez que no contexto do jornalismo digital, os “dados são arquivos de *bits* que no processamento de síntese, da junção de vários processos, tornam-se signos discretos ou contínuos que geram as semioses” (RAMOS, 2012, p.28). Eles não são informação, mas quando organizados em um formato, geram informação jornalística. “O formato cumpre a função de modelizar as linguagens potenciais que o signo informático pode gerar” (Ibidem, 2012, p. 28). Pensar no formato narrativo digital significa pensar em como os usuários se apropriam de serviços, tecnologias e dispositivos num contexto digital (SCOLARI, 2010, p.138). O *agenciamento coletivo interestratos* são as interligações que existem entre identificar e compreender em profundidade os estratos narrativos (BERTOCCHI, 2013). A partir deles é possível identificar quais nós foram amarrados, o que foi considerado importante para o sistema e o que eventualmente se revelou irrelevante em seu desenho.

A recolha de fontes, a seleção de informações e o formato da narrativa constituem critérios habitualmente ativados ao longo da modelagem narrativa não-digital e que são igualmente levados ao processo de modelagem da narrativa digital. No entanto, a consequência mais evidente da transposição das rotinas e da prática jornalística no momento do desenho do sistema narrativo é que a atuação do jornalista, uma vez realizada à moda tradicional, revela-se insuficiente e limitada — porque incide apenas em uma parte da modelagem narrativa, e não no sistema como um todo. Diríamos que incide muito mais no *frontend* (interface final) que no *backend* (bastidores). (BERTOCCHI, 2013, P. 59)

O jornalista não precisa ser multitarefa, mas precisa ter clareza de algumas práticas. No entanto, como sugere Bertocchi, no momento histórico em que vivemos o “jornalismo deve se abrir para ter contato com o *modus operandi* dos diversos estratos do sistema – o que lhe daria a vantagem de compreendê-lo e tirar deste entendimento algum proveito comunicacional” (BERTOCCHI, 2013, p. 60).

Assume-se que os jornalistas podem compreender sistemas sem precisarem se tornar matemáticos ou programadores, um ponto que já vem sendo trabalhado pela Universidade de Columbia através do curso de mestrado em Jornalismo e Computação, onde os alunos têm aulas de Introdução às bases de dados, Inteligência Artificial, Desenho da Interface do Usuário, Análise de Algoritmos e Sistemas de Visualização de Informação entre outras disciplinas atualmente ignoradas pelos cursos de Jornalismo tradicionais (BERTOCCHI, 2013, p. 60).

Para Daniela Osvald Ramos (2012), o profissional do jornalismo poderia atuar na modelagem da informação, dominando a técnica dos termos modelizados pelos formatos; usando as bases de dados, públicas ou privadas, como centro da criação e como fonte para reportagens assistidas por computador (RAC); aplicando a arquitetura de navegação; classificando conteúdos de forma semântica (tags, palavras-chave, *web* semântica), para que as informações contidas no formato possam ser recuperadas, tanto pela empresa quanto pelo usuário; buscando a clareza do que se quer de determinados formatos e a constituição das competências da equipe que o produzirá; realizando o *design* da interface na estética do formato; realizando a edição de áudio e vídeo, possuindo o domínio de *softwares* para a produção hipermídia. Tendo noções de *Flash*, com o uso de templates e das linguagens HTML, HTML 5, CSS (*Cascade Style Sheets*), PHP, noções de gerência de bancos de dados (MySQL, PHP Admin) e prática com publicadores de conteúdo, como *WordPress* e *Joomla*. Daniela Bertocchi (2014) complementa que é ainda importante a familiaridade do jornalista com o funcionamento da modelagem de dados (*dataentries*) para sistemas gerenciadores de conteúdos jornalísticos; otimização de textos, imagens e afins (um SEO); aplicação de algoritmos e inteligência artificial no jornalismo; curadoria de dados e informação; ser um jornalista que emita diversos tipos de informações *in loco*.

Escolhas e caminhos da modelagem jornalística

Na produção jornalística no ambiente online os profissionais podem escolher toda a estrutura da formatação narrativa (BERTOCCHI, 2013), o template, o design de informação, os fluxos de navegação, os metadados. Para Ramos (2011), Serres (2008) e Bertocchi (2013), o formato pode até ter um “formato-pai” onde, por exemplo, ocorra não a formatação da narrativa, mas o desenho do formato criando-se uma forma de antenarrativa, um formato onde as histórias x,y, z poderão ser trabalhadas de uma determinada forma. O que os autores colocam é que no ambiente digital, a narrativa permaneceu como uma extensão, um fazer da maneira clássica no ambiente digital, uma transposição dos gêneros clássicos para a plataforma digital. Isso aconteceu pelas razões já amplamente discutidas como a experiência prévia no meio impresso. No entanto, os formatos narrativos digitais são diferentes dos gêneros tradicionais. A primeira distinção é em relação a complexidade da narrativa. No jornalismo digital há inúmeras camadas e agenciamentos em relação as possibilidades de formatação narrativas (RAMOS, 2011). Os gêneros jornalísticos são estáveis e são dependentes da audiência. “(...) se o conceito de gênero surgiu a partir da língua natural, não é possível aplicá-lo diretamente às linguagens artificiais geradas por computadores no sentido de que são os gêneros que designam os formatos”. (RAMOS, 2012, p. 44).

De acordo com a professora do Programa de Pós-graduação em Meios e Processos Audiovisuais da Universidade de São Paulo Irene Machado (2007), o formato é o “design de gênero”¹⁶⁷, onde o formato não configura um gênero, mas o design de gêneros. Enquanto o gênero discursivo é da ordem da língua, o formato é da ordem das linguagens modelizadas pelos códigos culturais tecnológicos. Para a pesquisadora, não estando no ambiente semiótico da codificação tecnológica, o formato não pode ser concebido.

Os gêneros, ao serem redesenhados no contexto da mediação tecnológica, revelam a face metalinguística do formato: para criar linguagem é preciso processar linguagens e gêneros. Logo, o formato é uma noção que leva em conta todo um ambiente ecológico: a mídia (o sistema), os códigos (as linguagens) e as interações possíveis (as semioses). Fora dessas

¹⁶⁷ A pesquisadora Irene Machado afirmou no artigo *Gêneros e/ou formatos? Design de linguagem mediada* (2007), que o formato é a execução de design, o produto de um design e defende a noção de formato como “design de gênero”. (MACAHADO, 2007, p. 13)

disponibilidades de carácter ecológico, o formato não existe”
(MACHADO, 2007, p. 13).

Para Daniela Ramos (2011), o formato funciona como um artefato de produção jornalística, de interação, condicionada pelo dispositivo. É o resultado do desenho informático, o que pode conter muitas variáveis, pois a execução do design não é física, mas um processamento de códigos e linguagens na tela digital, uma síntese do design informático. É ao mesmo tempo um processo e um produto. No artigo *Gêneros e/ou formatos? Design de linguagem mediada* (2007), Irene Machado, defende que seja como um processo ou como um produto, os formatos possuem uma lógica interna em sua construção. Há invenções nos códigos culturais, mas também ocorre uma combinação de conhecimentos científicos e tecnológicos que resultam na criação de produtos. Elementos que juntos transformam a informação em mensagem.

Se, do ponto de vista dos códigos, esta transformação representa um processo modelizante, do ponto de vista dos conhecimentos é de design de informação que se trata. Considerando as implicações mútuas entre modelização e design, entre o processo e o produto, a ação de transformar a informação em mensagem cumpre um movimento de articulação de signos e produz diferentes semioses e significações. Neste funcionamento se explicita a lógica interna do sistema em suas diferentes mediações semióticas, isto é, a modelização e o design da informação (MACHADO, 2010, p. 97).

A ideia de múltiplos formatos, distribuídos em diferentes canais de comunicação e dispositivos se configura na atualidade como uma nova mentalidade para as empresas jornalísticas, pois apresenta para elas desafios nada simples de serem resolvidos, como a formação do jornalista, integração de redações e nova organização do fluxo de trabalho. Reforça que o sucesso do modelo de negócio das empresas jornalísticas passa por essa nova conceituação dos formatos narrativos: uma produção de informações da marca jornalística num fluxo contínuo independentemente do dispositivo (papel, papel digital, computador vestível ou outro qualquer), com inovação, sinergia, integração e muitas possibilidades em aplicativos para dispositivos móveis; ações que poderiam compensar as perdas de receita as quais vêm sofrendo nas últimas décadas.

Podemos dizer, portanto, que a experiência narrativa contemporânea sobre a qual procuramos aqui refletir é, em boa medida, costurada a partir de variadas interfaces gráficas acessíveis via diversos dispositivos que, juntas ou isoladamente, dão ao usuário o contato com as histórias do cotidiano escritas por jornalistas. Isso ocorre porque, no momento histórico em que vivemos, os computadores estão engastados no meio ambiente de forma visível ou invisível – o que tem sido denominado por computação ubíqua, internet das coisas, ambient intelligence, smart things, computação física, entre outros termos que dialogam entre si e tratam de um mesmo fenômeno visto de perspectivas computacionais, culturais ou sociais (RESMINI & ROSATI, 2011; GABRIEL; 2012 apud BERTOCCHI, 2013, p. 172)

Para Bertocchi é preciso pensar o formato além de um artefato, de uma entrega final, mas sim como processo, fluxo, resultado do sistema aberto e em constante adaptação. Assim, são identificados dois típicos formatos: aquele que se renderiza como uma visualização de dados (ou formato guiado por dados e metadados, o resultado do jornalismo de dados, o comportamento do sistema) e o formato a partir de algoritmo (criado por diversos atores humanos e não-humanos, como o algoritmo). Portanto, a observação do desenho do sistema narrativo revela três pontos importantes (BERTOCCHI, 2013, p. 208):

- 1) Dados, metadados e formatos têm a mesma importância no processo de composição da narrativa no jornalismo digital e ignorar uma dessas camadas pode representar uma perda comunicativa. Isso significa afirmar que o modelo assume que jornalistas desenvolvam competências e habilidades para selecionar, confeccionar, modelar, manipular e articular dados, metadados e formatos dentro de uma visão sistêmica do que seja narrar em meio digital;
- 2) Uma forma narrativa digital eficiente — do ponto de vista informacional, comunicacional, social, cultural e mesmo de modelo de negócio — tende a se concretizar mais naturalmente se o sistema narrativo tiver sido desenhado de antemão para suportar e fomentar tal eficiência;
- 3) Inexistem valores-notícia que estimulem o jornalista a trabalhar em camadas mais profundas da modelagem narrativa. Os conhecidos critérios de noticiabilidade não dão conta de estimular o jornalismo de dados ou a visualização de dados.

Analizamos que é nas camadas mais profundas e subterrâneas do contexto digital que são levadas a cabo decisões e aplicadas estratégias pouco familiares aos jornalistas. Já nas camadas mais superficiais encontram-se as decisões mais próximas à cultura jornalística tradicional: o desenho da

interface do usuário (a vitrine), a escolha de títulos, a disposição de imagens, as possibilidades de hiperligações de aprofundamento. (BERTOCCHI, 2013, p. 210).

Ao mencionar este desenho da interface, falamos do design. O que para nós é de grande relevância neste trabalho, uma vez que criar um design significa planejar, desenvolver critérios e elaborar um roteiro para chegar a algum ponto.

Design

O design funciona como um desenho, uma construção da escrita digital. No livro *O Mundo Codificado*, Villém Flusser (2007) aponta o design como central para a linguagem e que os jornalistas precisam ter consciência disso, de como funciona a máquina, o algoritmo, que é a codificação de uma série de instruções em uma linguagem. Onde acontece primeiramente o ato da estratégia a ser adotada, o desenho de um processo, onde há fluxos a serem rastreados e estes fluxos podem ser entendidos através do design. Isto se dá pela capacidade de atribuir modelos para as diversas ordens de códigos envolvidos no processamento informático (codificação, decodificação e transcodificação).

Para Ramos (2011), o design é um elemento estrutural do jornalismo de dados por ser exatamente o “processo de desenho que visa à elaboração de um artefato que no caso não é físico, mas sim as diferentes interfaces e formatos desenhados para organizar os dados” (RAMOS, 2011, p. 27). Não falamos aqui de dados primitivos, constantes ou variáveis, que são da ordem da informática e estão nos níveis de funcionamento estruturais do sistema. Falamos dos dados que são arquivos de bits, que são fotografias, vídeos, textos, animações, mapas, tabelas. Elementos que são passíveis de associações, comparações e que potencialmente geram visualizações. Dados que “em estado puro, não são informação, mas quando organizados em um formato, geram informação jornalística”. Sendo, portanto, os formatos interfaces para os bancos de dados. Por isso, conhecendo o processo de modelização da linguagem de um determinado texto da cultura, pode-se iniciar o processo de investigação de seus sistemas modelizantes ou conhecer a estrutura sob a qual o texto funciona.

A forma de descrição de como o sistema deve proceder é feita por formatos e visa alcançar uma maior eficiência e clareza na linguagem de modelagem. Como também estamos tratando de uma derivação da modelagem, a modelização do design informático, cujo resultante é o formato, precisamos também operacionalizar este processo. Assim, para criarmos formatos de informação jornalísticas digitais e a partir destes formatos, as narrativas, ao invés de escrevermos um texto escrito para encaixá-lo em um formato televisivo, radiofônico ou impresso, necessitamos projetar, desenhar, criar seu design. Este é o sentido com o qual usamos o termo “design”. A definição precisa do termo permite-nos articular o conceito, o qual utilizamos para a defesa de nossa tese (...). Assim podemos entender que o formato, modelizador das linguagens digitais, é modelizado pelo design informático que, por sua vez, é um encadeamento de modelizações de códigos de várias ordens, como vimos, e que o formato tem, por fim, vários níveis de modelização que são determinantes em sua apresentação (RAMOS, 2011, P. 46).

No curso sobre Visualização de Dados e Infografia, ministrado pelo jornalista e infografista Alberto Cairo (2014), após a apuração dos dados, é preciso fazer o esquema, como apresentar os dados. Este esboço, geralmente, é feito à mão. O que precisamos entender, são os processos. Não são só produtos, são fluxos, onde o jornalismo deve ser o centro da execução. Não é só a visualização, é o conceito envolvido que nos leva a entender o papel do design no ciberespaço. Há caminhos técnicos já traçados que auxiliam a entender melhor este universo; há técnicas sim, mas é preciso muito mais, é preciso também trazer humanidade aos dados. De acordo com Cairo (2014), fazer as perguntas certas e apresentá-las da melhor forma possível.

É possível decidir objetivamente sobre quais formas e gráficos são mais adequados? 1. Pense sobre a audiência e sobre a publicação; 2. Pense nas perguntas que seu gráfico deve ajudar os leitores a responder; 3. Você pode entendê-lo sem ter que ler cada figura? (CAIRO, 2014, tradução nossa)¹⁶⁸

O doutor em Marketing pela Universidade da Virgínia, Andrew Abela, desenvolveu alguns métodos para o desenvolvimento de apresentações de dados complexos e informações com gráficos, principalmente voltado para empresas e baseado em vários estudos que incluíram também diferentes áreas de conhecimento, como psicologia, marketing, comunicação e design. O diagrama desenvolvido por Abela, chamado

¹⁶⁸ “Is it possible to objectively decide which graphic forms are more appropriate? 1. Think about the audience and the publication; 2. Think of the questions your graphic should help readers answer; 3. Can you understand it without reading every single figure?” (CAIRO, 2014). Slide referente a segunda semana de aula do curso de Visualização de Dados e Infografia, ministrado pelo prof. Alberto Cairo em abril de 2014.

Chart Suggestions – A Thought-Starter (em português *Sugestões de Gráficos – Uma ideia inicial*), por exemplo, serve como um guia para selecionar um gráfico apropriado. Este guia é muito popular porque agrupa gráficos por função: comparação, distribuição, composição e relacionamento. Ele é baseado em extensa pesquisa empírica e já foi testado entre diversas empresas como a Microsoft, a Exxon-Mobil, a Kimberly-Clark, eBay e Motorola.

Figura III. 12 – Versão do diagrama Chart Suggestions – A Thought-Starter, de Andrew Abela

Fonte: TEIXEIRA, 2014, p. 137.

Como exemplo de uso deste guia, podemos citar a vertente relação. Caso os dados possuam duas variáveis para serem mostradas ao público, é mais indicado o uso de um gráfico de dispersão e caso possua três ou mais variáveis seria um gráfico de bolhas. Há repleta bibliografia na área; para muitos pesquisadores¹⁶⁹ este próprio guia chega a ser um pouco confuso e apresenta algumas falhas, sendo relevante leituras

¹⁶⁹ A exemplo de FEW, Stefan (2015). Blog Stephen Few - Abela's Folly – A Thought Confuser. Disponível em: <https://www.perceptualedge.com/blog/?p=2080> - Abela's Folly – A Thought Confuser. Acesso em 20 de janeiro de 2016.

mais aprofundadas nesta área do *design thinking*¹⁷⁰ e *visual thinking* (pensamento visual)¹⁷¹ que servem como base para a matriz de construção de infográficos. A simples escolha de uma cor ou o formato de um gráfico de barras devem ser guiados pelos estudos já empreendidos na área da percepção visual, uma vez que determinados elementos visuais possuem resultados já identificados por serem mais indicados para atender com propriedade as inclinações e “necessidades” de nosso cérebro. Nós temos a tendência natural do cérebro em buscar equilíbrio e harmonia nas coisas ao redor e também em procurar por uma continuidade coerente. Não é a toa que formas com aparência distorcida podem provocar desconforto visual. Apesar de não ser o foco desta pesquisa, como meio de contextualização do processo de visualização de dados, vamos nos estender em elementos da visualização da informação para a compreensão de etapas que muitas vezes não aparecem para quem apenas vê o resultado final. São muitas idas e vindas dentro do sistema narrativo para que o resultado obtido seja satisfatório. Diversas adaptações são necessárias nas peças jornalísticas. O design é algo adaptável e mutável de acordo com as condições no qual é exposto e a ferramenta que será proposta segue essa mesma lógica.

Há elementos estruturantes que participam da construção infográfica. Ainda mais quando falamos de peças que advêm da visualização de dados e que demandam conhecimentos, como a programação em banco de dados, que não estão relacionados diretamente à área de design, mas que podem resultar em um nível de profundidade e contribuição mais amplos. Para nós, o que é mais relevante nesta pesquisa é o que defendem alguns autores da área, como Edward Tufte e Alberto Cairo. Eles apontam que os infográficos devem priorizar elementos visuais que sejam de fato funcionais e que contribuam para a eficácia do entendimento do assunto abordado, não sendo apenas ornamentais. Ao abordar o pensamento visual, o autor Dan Roam (2012) discute pilares que são os principais elementos presentes no processo de desenvolvimento de infográficos: 1ª etapa: olhar ou coletar dados; 2ª etapa: ver ou selecionar e criar padrões com dados; 3ª etapa: imaginar ou estabelecer conexões não aparentes; 4ª etapa: mostrar ou traduzir visualmente as etapas anteriores com a

¹⁷⁰ De acordo com a Wikipédia, *Design Thinking* é o conjunto de métodos e processos para abordar problemas, relacionados à aquisição de informações, análise de conhecimento e propostas de soluções. Disponível em: https://pt.wikipedia.org/wiki/Design_thinking. Acesso em: 5 de abril de 2016.

¹⁷¹ Dan Roam, autor do livro *The Back of the Napkin: Solving Problems and Selling Ideas with Pictures* defende a ideia de que é possível resolver problemas através da estruturação de estratégias com desenhos. O pensamento visual trata exatamente destas modelagens aplicadas a negócios. Uma vez que um desenho pode facilitar a compreensão.

escolha de recursos gráficos. A pesquisadora Carla Cristina da Costa Teixeira (2014), em sua tese doutoral no Departamento de Artes & Design, na Pontifícia Universidade Católica do Rio de Janeiro aplicou com sucesso os conceitos de Dan Roam na construção de infográficos. Para ela, responder as perguntas propostas por Roam ajudam a colocar o profissional do design no lugar do receptor:

(...) as perguntas também representam um exercício para tentar captar que perguntas o leitor gostaria que fossem respondidas pelo infográfico, ajudando o designer a se colocar na posição do receptor da mensagem visual, e não apenas como mero emissor sem levar em consideração as necessidades e peculiaridades dos indivíduos para quem está se comunicando, como faz um verdadeiro design thinker. (TEIXEIRA, 2014, p. 162).

Para Roam, alguns designers têm o hábito de suprimir a etapa do olhar e coleta aos dados ou fazê-los, muitas vezes, de forma superficial “Compreender o pensamento visual como um processo pleno significa que o ponto de partida não é aprender a desenhar melhor; é aprender a olhar mais atentamente” (ROAM, 2012, p.46). O nível de profundidade da pesquisa de dados é fundamental para a criação de boas peças, cruzar dados e dar sentido a eles de forma visual. Esta fase do olhar tem um caráter imersivo sobre determinado assunto, é de cunho investigativo e muitas vezes inclui outros meios, como entrevistas, depoimentos, elementos que venham a contribuir para o conhecimento. As etapas que se desenvolvem posteriormente são relevantes na filtragem de informações que serão apresentadas. Criações de padrões e organizações lógicas. Roam parte inclusive das perguntas-chave que nos lembram o lead jornalístico: quem; o quê, quanto, onde, quando, como e porquê.

As demais etapas propostas por Roam remetem a apresentação dos conteúdos, de imaginar a apresentação dos conteúdos. “Imaginar é a maneira como deixamos os olhos de nossa mente assumirem o comando de modo que possamos ver coisas que não estão fisicamente visíveis” (ROAM, 2012, p.90). É onde a criatividade floresce, que diante de um grande volume de dados, permite diversos caminhos. São feitas adaptações, escolhidas construções visuais de acordo com circunstâncias e contextos diversos.

Roan chega inclusive a sugerir “seis maneiras de ver” os dados e apresentá-los, no estilo qual é a pergunta e qual recurso gráfico melhor pode correspondê-la (ROAM, 2012, p.128):

1. Quem/o quê? – representação qualitativa – retrato/ilustração;
2. Quanto? – representação quantitativa – gráfico;
3. Onde? – posição no espaço – mapa;
4. Quando? – posição no tempo – linha do tempo;
5. Como? – causa e efeito – fluxograma;
6. Por quê? – dedução e previsão - gráfico de variáveis.

Carla Costa Teixeira, com base nos trabalhos de Roam, chegou a desenvolver uma matriz de construção de um infográfico:

Figura III. 13 – Matriz de Construção Infográfica (TEIXEIRA, 2014)

Fonte: TEIXEIRA, Carla, 2014, p. 182.

São cinco as etapas da matriz de construção de infográficos propostas por Teixeira (2014, p. 179-182):

- 1 - Primeira etapa – Definição – Essa etapa se refere à definição precisa do tema em questão que será abordado pelo infográfico;
- 2 - Segunda etapa – Pesquisa de dados – Nesse momento o designer deve se abastecer do máximo de dados possível em relação ao tema definido na etapa anterior;
- 3 – Terceira etapa – Seleção de dados – Essa fase se refere ao processo de selecionar os dados que foram coletados para a construção do infográfico, de acordo com uma série de critérios que devem ser estabelecidos pelo designer e também seguindo técnicas de seleção já mencionadas na pesquisa;
- 4 – Quarta etapa – Estudos de layout – A ferramenta sugerida para começar a pensar como traduzir visualmente os dados selecionados para o infográfico é o SQVID¹⁷², que aciona o modo imaginar presente no método proposto por Dan Roam. Além disso, estimula o designer a realizar conexões entre os dados que a princípio não eram aparentes, auxiliando no processo de descobrir, aprofundar e elucidar questões para o público-alvo da peça gráfica;
- 5 – Quinta etapa – Implementação - A etapa final da matriz de construção é definida pela escolha de um dos estudos de layout ou então a partir do cruzamento de mais de um deles.

Concordando com Ramos (2011), acreditamos que os aspectos do formato do design digital não se restringem à área de navegação na tela, à própria navegação ou a rolagem. “O formato é resultado do desenho informático, o que pode conter muitas variáveis, pois a execução do design não é física e palpável, mas um processamento de códigos e linguagens na tela digital, uma síntese do design informático”(RAMOS, 2011, p. 45). Os formatos estão armazenados em bancos de dados de empresas jornalísticas e acessados pelos usuários que solicitam o formato por meio de um clique de mouse em interfaces variáveis.

¹⁷² A sigla SQVID foi elaborada por Roam (2012, p. 98) e corresponde a primeira letra de atributos a serem respondidos por cada uma das perguntas: S (simples ou elaborado), Q (qualidade ou quantidade), V (visão ou execução), I (atributos individuais ou comparação) e D (mudança ou status quo). Foram escolhas do autor para facilitar a leitura e memorização destes atributos.

Certamente, os dispositivos de acesso podem mudar nos próximos anos, a exemplo do iPhone, da Apple, cuja interface é manipulada através do toque na tela, assim como o iPad, lançado em 2010, todos com uma tecnologia proprietária, já que alguns movimentos que fazemos para acessar o touchscreen destes dispositivos são patenteados pela empresa. Neste caso, as particularidades de funcionamento dos dispositivos condicionam a geração dos formatos (RAMOS, 2011, p. 47).

De acordo com o pesquisador Eduard Tufte, no livro *The Visual Display of Quantitative Information* (1983), há muito *chartjunk*, ou seja, lixo gráfico. O autor considera que esses elementos apenas contribuem para poluir visualmente a peça gráfica e tirar a atenção do que seria realmente importante em termos de conteúdo. Fato que também foi evidenciado na pesquisa desenvolvida por Giannella (2014), na análise de 270 peças gráficas de infográficos. De acordo com ela, as ferramentas online e gratuitas avançam em termos de armazenamento e organização de dados, agilidade e baixo custo de produção das peças; mas por outro, podem ser limitadas em termos de raciocínio e imaginação. Os problemas também giram em torno da incompatibilidade do projeto gráfico do veículo, da compartimentação das informações em blocos e do modo de organização não condizentes com a natureza dos dados:

- incompatibilidade ou contraste com o projeto gráfico do veículo de comunicação em termos de: escala de cor, tipografia, diagramação dos elementos visuais e textuais;
- compartimentação da informação em blocos: criação de duas ou mais representações gráficas com diferentes títulos, dispostas verticalmente na página, fragmentando o conteúdo e dificultando a relação entre os dados visualizados;
- modo de organização da informação não condizente com a natureza e relação dos dados que deveriam ser evidenciados. (GIANNELLA, 2014, p. 127).

A criação de programas e algoritmos que substituem o trabalho humano de reportagem envolve uma série de decisões que devem ser passíveis de explicação e responsabilização para todos os afetados. Aplicativos de *softwares* baseados na *web* precisam da mediação humana, profissional, para emergir e interpretar as informações. Não basta torná-las acessíveis ou legíveis; é preciso torná-las compreensíveis e relevantes, ter valor jornalístico.

Valor na contemporaneidade

Como já apontado no início deste capítulo (página 102), concordamos com ANDERSON, BELL & SHIRKY (2012), com o fato de que jornalista foi deslocado na cadeia editorial e hoje não produz informações iniciais. O profissional exerce a função de verificar, interpretar e dar sentido à enxurrada de textos, áudios, fotos e vídeos produzidos pelo público e/ou que são de interesse e necessidade de conhecimento do público. No relatório do *Tow Center for Digital Journalism da Columbia Journalism School*, os pesquisadores reforçam a ideia de que neste novo contexto cultural, o jornalismo é afetado em três grandes áreas: no fazer jornalismo, na instituição jornalística e no ecossistema de mídias. Por exemplo, TV e redes sociais, uma das dinâmicas é a presença da segunda tela. Entre os apontamentos para os jornalistas tanto do mercado formal, quanto do informal teremos a credibilidade do jornalista cada vez mais no nome do profissional e um jornalismo mais militante, opinativo. Entre as habilidades que o contexto cada vez mais vai pedir estão narrativas e novas alfabetizações jornalísticas. A originalidade no raciocínio, que deve provocar mudanças, incitar a ação e o carisma, uma vez que “gente segue gente”, ainda mais em um mundo de redes onde as conexões estão facilitadas.

O jornalismo é uma ferramenta, mas não somente técnica. As questões envolvem contextualizações e a prática profissional deve estar associada ao contexto social e cultural. Jornalismo não é informação primária e neste sentido a apuração é a ferramenta fundamental que justifica a presença da profissão, que só terá o desenvolvimento da utilidade do conteúdo com a compreensão das relações de poder, as traduções e desvendamentos destas relações no âmbito social. Nesse novo ecossistema torna-se relevante entender o que novos atores podem fazer hoje melhor que o jornalismo de modelos anteriores? Que papel o jornalista pode desempenhar melhor do que ninguém? O que os novos atores, *softwares* fazem melhor que os jornalistas? Para a pesquisadora Daniela Osvald Ramos (2015, anotações)¹⁷³ é preciso pensar o jornalismo fora do jornalismo. Quanto mais esse processo acontecer, mais divertido, mais produtivo, mais criativo o jornalismo será. “Devemos encarar a experimentação voltada à inovação como algo a praticar e não simplesmente tolerar”,

¹⁷³ Em aula realizada na Universidade de São Paulo no dia 27 de abril de 2015, disciplina Jornalismo Online.

disse. “Precisamos também bancar a humanidade, trazer as pessoas para os dados, complementou Ramos.

Em meio as mudanças e as adaptações de novas formas de contar histórias baseadas em dados aparecem novas possibilidades para o jornalismo. No entanto, elas exigem diferentes formas de organização e, como já mencionado, somente incorporar técnicas não serão suficientes para a adaptação a este ecossistema. Torna-se importante refletir sobre a formação dos jornalistas nas universidades brasileiras sob a perspectiva das habilidades e competências necessárias para o fomento do pensamento computacional no jornalismo.

(...) talvez não seja possível competir com o mercado informativo atual se não se conta com especialistas em reunir, organizar e interpretar números; ou, em outras palavras, em dá-los sentido e apresentá-los visualmente de maneira que possam ser aprendidos e estudados (CAIRO, 2011, tradução nossa)¹⁷⁴

O jornalista é potencialmente um designer de experiência. Ele não apenas "escreve o texto", mas é a figura também capaz de modelar a narrativa em camadas, com equipes humanas e robôs, tendo como objetivo uma experiência narrativa centrada nos usuários. Se familiarizado com as camadas mais subterrâneas do sistema narrativo, o jornalista poderia atuar mais no remodelamento do sistema, encontrando e criando oportunidades de melhor comunicar suas histórias no ciberespaço. O formato visualização de dados tem processos próprios e estes processos dependem dos dados, das linguagens, das ferramentas, assim como dos locais de armazenamento e formas de resgate.

Considerações

Neste capítulo sustentamos que o jornalismo de dados e a visualização de dados estão inseridos em sistemas narrativos condizentes com processos inovadores no jornalismo contemporâneo, onde o jornalista não é somente um narrador de fatos, ele é o

¹⁷⁴ “Tal vez no sea posible competir en el mercado informativo actual si no se cuenta con especialistas en reunir, organizar e interpretar números; o, en otras palabras, en darles sentido y presentarlos visualmente de manera que puedan ser aprehendidos y estudiados” (CAIRO, 2011). Disponível em: <http://blogs.elpais.com/periodismo-con-futuro/2011/04/periodismo-de-precision-visualizacion-datos.html>. Acesso em Visualizado em 15 de abril de 2014

responsável por coletar, apurar, interpretar e contextualizar as informações. Recorre a gráficos interativos, solicita imagens e informação ao público/leitor, disponibiliza acesso direto a bancos de dados e distribui matérias por redes sociais.

Defendemos que o jornalista não foi substituído e muito menos extinto, mas sim, neste cenário, exerce funções de verificar, interpretar e dar sentido à enorme quantidade de textos, áudios, fotos, vídeos em narrativas originais e criativas. É um curador de notícias que se familiarizado com as camadas subterrâneas do sistema narrativo encontra melhores maneiras de contar histórias. Neste sistema narrativo atuam diferentes atores humanos e não-humanos, em construções iniciadas em codificações maquínicas, com bases de dados que dialogam com outras camadas até chegar a uma interface com o usuário. O usuário, por sua vez, dá continuidade ao fluxo informativo comentando, compartilhando, se apropriando e resignificando o conteúdo a que teve acesso.

Novas possibilidades para o jornalismo levam a diferentes formas de organização e somente a incorporação de técnicas não serão suficientes para a adaptação a este novo ecossistema. É preciso dar valor e funcionalidade ao jornalismo retomando perguntas fundamentais para o exercício da profissão: o que é possível fazer hoje melhor do que o jornalismo de modelos anteriores? Que papel o jornalista desempenha melhor do que ninguém? O que os novos atores, *softwares*, fazem melhor que os jornalistas? Onde, neste processo, os jornalistas fazem a diferença?

A produção jornalística no ambiente online é complexa, com inúmeras possibilidades e camadas de formatação narrativa. O design passa a ser um elemento estrutural no jornalismo de dados, uma vez que é o responsável pelo desenho, pela a organização dos dados em formatos, sendo, portanto, estratégico. Os profissionais envolvidos podem escolher toda a estrutura: template, design, navegação, metadados. Pode até existir um “formato-pai”, mas está longe de ser a maneira clássica do jornalismo. Os formatos narrativos digitais são diferentes dos gêneros tradicionais. Enquanto o gênero discursivo é da ordem da língua, o formato é da ordem das linguagens modelizadas pelos códigos culturais tecnológicos.

A complexidade contemporânea também é refletida nos estudos da área, uma vez que são necessários diferentes olhares e reflexões teórico-metodológicas para entender os fenômenos. No quarto capítulo se encontram os caminhos metodológicos percorridos nesta pesquisa e os rumos para as observações de campo que serão apresentadas no capítulo V.

IV. CONSTRUINDO A PARTIR DA DESCONSTRUÇÃO – CONSIDERAÇÕES METODOLÓGICAS

Resumo: O que é a visualização de dados no jornalismo? Como ela se insere na prática e na narrativa do jornalismo de dados proporcionando diferenciais relevantes? Este capítulo tem por objetivo apontar os caminhos metodológicos percorridos para tentar resolver os questionamentos deste estudo. Foi adotada uma pesquisa multimetodológica tendo os pressupostos da centralidade, transversalidade e resiliência na busca de diferentes olhares a partir dos dados.

Palavras-chave: Pesquisa multimetodológica; Pesquisa Empírica em Comunicação; Estudo de Caso; Grounded Theory

"Embora seja fácil mentir com estatísticas, é ainda mais fácil de mentir sem elas." (Frase atribuída ao estatístico americano Frederick Mosteller, presente na obra de Charles Murray, 2005, tradução nossa)¹⁷⁵

Diferentes olhares

Foram inúmeros os desafios desta pesquisa, não menos complexo o de como trabalhar a metodologia e análise dos dados. Um assunto dinâmico, ainda com poucas referências, exigiu de nós um olhar multidisciplinar, que confrontou teorias e métodos de investigação. O texto *Centralidade, transversalidade e resiliência: reflexões sobre as três condições da contemporaneidade digital e a epistemologia da Comunicação*, da pesquisadora Elizabeth Saad (2015), veio nos auxiliar e ampliar horizontes sobre como posicionar nosso estudo no cenário contemporâneo. Ele sintetiza as características e, por consequência, os desafios dos estudos em Ciências da Comunicação envolvendo as temáticas da mídia digital, cibercultura, mídias sociais,

¹⁷⁵ "While it is easy to lie with statistics, it is even easier to lie without them" (Charles Murray). Disponível em: http://www.jstor.org/stable/20061179?seq=1#page_scan_tab_contents. Acesso em 16 de março de 2016.

do cenário contemporâneo que exige diferentes reflexões teórico-metodológicas para entender os fenômenos da área.

Partindo de uma centralidade da comunicação, concordamos com Saad que não podemos mais distinguir um espaço específico para o estudo do digital dentro da Comunicação. As Tecnologias Digitais de Informação e Comunicação atravessam, transversalmente, todas as atividades comunicativas e informativas sejam elas analógicas e/ou off-line, o que requer resiliências em relação à estruturação das abordagens teórico-metodológicas que sustentam as pesquisas e reflexões do campo. As fronteiras do campo de estudos passam por processos de dissolução diante dos diversos “paradigmas-teorias-modelos-metodologias vinculados à delimitação da Comunicação como campo de estudo da transmissão de mensagens por meio de dispositivos de mídia” (SAAD, 2015, p.3). Há uma maior complexidade dos estudos de Comunicação no estreitamento com as Tecnologias Digitais de Informação e Comunicação e a necessidade de convivência com reinterpretações dos fenômenos como interatividade, midiaticização, virtualidade.

A diversidade de plataformas midiáticas traz o desafio de conceituações e operacionalizações de múltiplas mídias. Ao abordar a complexidade, nos remetemos ao que Edgar Morin (2006) vem discutindo e com a proposta de uma visão mais ampla percebemos que a transversalidade, como apontado por Saad, passa a ser uma capilaridade das tecnologias digitais, de atuação simultânea:

A transversalidade é capilaridade das tecnologias digitais atuando simultaneamente nos processos que operam as atividades comunicativas, nos sistemas que integram processos anteriormente fragmentados, nos dispositivos cada vez mais convergentes devido às affordances que incorporam funções de mobilidade e geolocalização, interatividade aos suportes comunicativos clássicos, e nos próprios produtos midiáticos (SAAD, 2015, p. 6).

Tendo a centralidade da Comunicação no tecido social e a presença do digital de forma transversal nos processos, produtos e práxis do campo, Saad aponta a resiliência como condição para a estruturação epistemológica. Uma resiliência benéfica na possibilidade de incorporação de teorias/modelos/metodologias de campos correlatos à Comunicação e que por consequência facilitam outros saberes.

Entendemos aqui por resiliência a capacidade de um sistema ou uma organização se antecipar e se adaptar a rupturas, eventos, lidar com as mudanças e reconstruir seus valores e estruturas a partir destes movimentos. Numa visada objetiva e sem qualquer caráter de crítica, assistimos hoje a um processo de busca de diferentes vertentes teóricas, multiplicidade de autores e propostas metodológicas para a sustentação epistemológica de estudos e pesquisas que envolvam a questão digital na Comunicação, seja ela vista como um elemento transversal nos processos ou nos produtos, seja como elemento central de discussão (SAAD, 2015, p.9).

Para nós, a metodologia deve ser analisada não somente como técnica de pesquisa, mas como a explicitação da construção do objeto e dos aspectos metateóricos, epistemológicos da pesquisa. Em suma, refletem também as escolhas do pesquisador. Trata-se, portanto, da percepção de um sujeito que sofre influências dos condicionantes sociais. É preciso pesquisar com seriedade e consciente das dificuldades do trajeto. Neste projeto, viu-se os fenômenos da comunicação digital como manifestações da comunicação humana, onde os conteúdos, que ocorrem no ambiente tecnológico dos *bits* e que utilizam ferramentas técnicas possibilitam promover trocas, interações, relações de sociabilidade.

Diante da complexidade da temática na contemporaneidade foi adotada uma estratégia multimetodológica envolvendo revisão bibliográfica, observação não participante orientada pela Pesquisa Empírica em Comunicação (LOPES, 2010), estudo de caso (YIN, 2001) do blog Estadão Dados, do jornal O Estado de S. Paulo e, posteriormente, pressupostos da Grounded Theory (TAROZZI, 2011) para o olhar mais detalhado sobre o fenômeno estudado.

O esquema da figura a seguir mostra o caminho percorrido para a familiarização com o tema até a aplicação dos princípios da Grounded Theory (GT). Princípios, porque a proposta deste estudo não foi a busca por uma teoria, mas obter informações a partir de diferentes olhares e deixar também que os dados falassem por si, questões que explicaremos melhor ao longo deste capítulo.

Figura IV. 14 – Familiarização com o tema visualização de dados

Fonte: Mayanna Estevanim, 2016.

As escolhas aqui se deram pela necessidade de cercar o objeto e tentar compreendê-lo por diversos pontos de vista, principalmente por estar em uma área inovadora de conhecimento e transdisciplinar; em uma zona fronteira onde o consumo e a

produção de conteúdos apresentam distinções em relação a práticas de outros momentos históricos; onde os recursos tecnológicos possibilitam novos fazeres jornalísticos, novas formas de interação. Além da revisão da literatura e das disciplinas cursadas¹⁷⁶, diversas atividades complementaram os estudos. Foram artigos publicados¹⁷⁷, observação empírica em sites nacionais, MOOC's¹⁷⁸, PAE¹⁷⁹, congressos e discussões no grupo de pesquisa COM+¹⁸⁰. Buscamos através do caminho metodológico saber como o jornalismo de dados, em âmbito nacional, vem utilizando a visualização de dados. O que é a visualização de dados no jornalismo e como ela se insere na prática e na narrativa do jornalismo de dados proporcionando diferenciais relevantes.

ETAPA - pesquisa bibliográfica

Levantamento bibliográfico que permitiu compreender melhor os estudos do jornalismo digital em base de dados, da visualização de dados e do jornalismo pós-industrial. De acordo com a professora Maria Immacolata Vassalo de Lopes (LOPES, 2010), além do universo teórico feito a partir do paradigma científico, a pesquisa bibliográfica visa situar o problema em relação às pesquisas existentes, mesmo aquelas de orientações teóricas diferentes. Visa descrever o “estado de conhecimento” sobre o objeto, o que já foi estudado sobre ele.

¹⁷⁶ 1- Metodologia da Pesquisa em Comunicação; 2- Teorias da Produção Partilhada do Conhecimento em Comunicação Digital: Comunidade e Universidade; 3- Mídias Sociais Audiovisuais e o Espaço Público Contemporâneo: Práticas e Processos em Mutação; 4- Comunicação e Redes na Era Digital: Pensamento Contemporâneo e Tendências Tecnológicas; 5- Epistemologias Reticulares: Questões Sobre o Método no Estudo das Redes Digitais e Perspectiva de Conhecimento Através da Interação com Computadores.

¹⁷⁷ O artigo *Visualização de Dados Complexos no Jornalismo Digital: uma reflexão sobre os conceitos de dispositivo* foi publicado nos anais do Intercom Nacional, GP Teoria do Jornalismo, XIV Encontro dos Grupos de Pesquisa em Comunicação, evento componente do XXXVII Congresso Brasileiro de Ciências da Comunicação realizado em Foz do Iguaçu (PR), entre 2 e 5 de setembro de 2014. O artigo *Jornalisticar - Expertises na Visualização e no Jornalismo de Dados no Digital*, será publicado nos anais do VIII Simpósio Nacional da ABCiber, realizado entre 3 e 5 de dezembro de 2014, na Escola Superior de Propaganda e Marketing (ESPM), em São Paulo. O artigo *Formatos digitais e qualidade no jornalismo brasileiro em base de dados*, de autoria conjunta com as professoras Daniela Osvald Ramos, da Universidade de São Paulo e Egle Müller Spinelli, da Universidade Anhembi Morumbi foi publicado nos anais do IV Congresso Internacional de Ciberjornalismo, realizado nos dias 4 e 5 de dezembro de 2014 na Faculdade de Letras da Universidade do Porto. O artigo *Visualização de dados complexos no jornalismo digital* entrará nos anais da V Jornada Acadêmica Discente PPGCOM USP, realizado em 10 de dezembro de 2014, na Escola de Comunicações e Artes, da Universidade de São Paulo. O artigo *Visualização de dados complexos no jornalismo e os estudos de redes sociais digitais na contemporaneidade* entrará nos anais do XIV Congresso Ibero-americano de Comunicação – IBERCOM 2015, realizado entre 29 de março e 02 de abril de 2015, na Escola de Comunicações e Artes, da Universidade de São Paulo.

¹⁷⁸ Realização do curso massivo e aberto *Visualização de Dados e Infografia*, da University of Texas at Austin em parceria com o Knight Center Journalism in the Americas, entre os dias 14 de abril e 12 de maio de 2014, ministrado pelo jornalista e professor da School of Communication of the University of Miami, Alberto Cairo. Realização do curso massivo e aberto *Herramientas Digitales para el Periodismo de Datos*, da University of Texas at Austin em parceria com o Knight Center Journalism in the Americas, entre os dias 18 de agosto e 21 de setembro de 2014, ministrado pela jornalista e professora Sandra Crucianelli. Curso *Math for Journalists Made Easy: Understanding and Using Numbers and Statistics*, entre os dias 1º e 28 de junho de 2015, realizado pela University of Texas at Austin. Curso *Lógica de programação*, realizado em 1º de julho de 2015, pela Softblue.

¹⁷⁹ O estágio no Programa de Aperfeiçoamento do Ensino foi realizado no 2º semestre de 2014 e no 1º semestre de 2015, ambos na disciplina Jornalismo On-line, Escola de Comunicação e Artes.

¹⁸⁰ <http://grupo-ecausp.com/2010/05/grupo-pesquisa-commais/>

Acumula-se uma documentação sobre o objeto que fornece o estado atual do conhecimento sobre o tema de pesquisa, isto é, “o que se diz dele”. Esse levantamento bibliográfico, quando bem realizado, oferece textos de apoio a todas as dimensões da pesquisa, envolvendo questões epistemológicas, teóricas, metodológicas e técnicas” (LOPES, 2010, p. 148).

2. ETAPA – Definição inicial das categorias de observação e busca dos veículos de comunicação para análise

Primeiramente foram buscados critérios que indicassem que o conteúdo fazia parte do Jornalismo Digital de Base de Dados. Em um segundo momento critérios que determinassem que se tratava de uma visualização de dados e em um último momento, diante do avanço das observações do estudo, foram definidos os critérios de observação envolvidos no processo da visualização de dados, para que a partir destes apontamentos fosse possível chegar ao veículo que seria fruto do estudo. Vamos a eles:

a) O JDBD prevê funções para as bases de dados que, em sua maioria, se relacionam com o sistema narrativo. Com base no trabalho iniciado por Susana Barbosa (2008), a pesquisadora Daniela Bertocchi (2013) fez um recorte de quais funções impactam mais diretamente no sistema narrativo:

Indexar e classificar as peças informativas e os objetos multimídia; - Integrar os processos de apuração, composição e edição dos conteúdos; - Conformar padrões novos para a construção das peças informativas; - Propiciar categorias diferenciadas para a classificação externa dos conteúdos; - Estocar o material produzido e preservar os arquivos (memória), assegurando o processo de recuperação das informações; - Permitir usos e concepções diferenciados para o material de arquivo; - Garantir a flexibilidade combinatória e o relacionamento entre os conteúdos; - Gerar resumos de notícias estruturados e/ou matérias de modo automatizado; - Armazenar anotações semânticas sobre os conteúdos inseridos; - Habilitar o uso de metadados para análise de informações e extração de conhecimento, seja por meio de técnicas estatísticas ou métodos de visualização e exploração como o *data mining*; - Orientar e apoiar o processo de apuração, coleta e contextualização dos conteúdos; - Regular o sistema de categorização de fontes jornalísticas; - Sistematizar a identificação dos profissionais da Redação; - Cartografar o perfil dos usuários; - Transmitir e gerar informação para dispositivos móveis (celulares, computadores de mão, entre outros). (BERTOCCHI, 2013, p. 33).

Como forma de delimitação, partimos das categorias descritivas adotadas por BARBOSA (2008) no JDBD: dinamicidade; automatização; inter-relacionamento e hiperlinkagem; flexibilidade; densidade informativa; diversidade temática e visualização. Para Bertochi (2013), a visualização se manifesta no momento de narração (formatação narrativa), enquanto que os demais critérios podem ser observados no momento de antenarrativa. A visualização de dados, no entanto, possui em sua sistemática um ir e vir dentro do sistema narrativo (fato mencionado no capítulo III).

b) A partir destas primeiras categorias foram adotados critérios mais específicos de observação. De acordo com Cairo (2011), a visualização de dados é a disciplina que se ocupa de criar ferramentas visuais, as vezes interativas, para que uma determinada audiência, geralmente composta por pesquisadores e acadêmicos, possam estudar e explorar os dados. Em uma visualização o usuário ocupa o lugar de editor elegendo quais dados quer ver e como quer vê-los.

A forma de apresentação e de exploração são funções que caracterizam e separam a infografia da visualização de dados como conteúdo final apresentado ao leitor. Para tanto, como forma de delimitação, observamos quais conteúdos agregavam grande volume de dados e permitiam que o usuário construísse suas narrativas e explorasse um grande volume de dados de forma visual.

3. ETAPA – Coleta de dados para contextualização

a) Paralelamente à segunda etapa foram levantadas iniciativas que tiveram como foco a visualização de dados e o jornalismo: congressos, premiações, blogs, cursos online gratuitos na área de visualização, buscas para contextualizar a prática em âmbito nacional e internacional.

b) Observação não-participante de iniciativas jornalísticas em visualização de dados em veículos de grande mídia nacionais e internacionais. O processo de observação aqui descrito faz parte da segunda fase da pesquisa empírica em Comunicação

(LOPES, 2010), onde o trabalho passou a adquirir mais corpo diante dos fenômenos acompanhados.

A observação é o momento da ruptura epistemológica na prática e, como o processo de construção dos dados da pesquisa, abrange duas operações: amostragem e coleta dos dados. Para atender a proposta deste estudo foi adotada uma combinação de observações diretas e indiretas, a partir de entrevistas, observação não-participante e participações em cursos, palestras e grupos de discussão com temáticas correlacionáveis. Tendo em vista que as técnicas de coleta escolhidas dão significado epistemológico ao tratamento do objeto e também teórico aos problemas que se apresentam a este objeto, foram adotadas prioritariamente as de amostragem não-probabilística, de dados qualitativos, uma vez que “na amostra não probabilística a inferência estatística não pode ser legítima, uma vez que o pesquisador não conhece a probabilidade que cada unidade tem de ser selecionada para fazer parte da amostra.” (LOPES, 2010, p. 145). Abordamos um fenômeno recente, ainda em construção e desenvolvimento. Não temos, portanto, fundamentos para generalizar situações ou condicionar os fenômenos de alguma forma.

4. ETAPA – estudo de caso

A partir dos referenciais teóricos adotados e do paradigma Jornalismo Digital de Base de Dados foi selecionado o blog Estadão Dados, do jornal O Estado de S. Paulo, para o estudo de caso. A escolha se deu por ter no blog tanto conteúdos gráficos, quanto projetos em visualização, pelas constantes atualizações de conteúdo e por estar em um cenário brasileiro de investimentos nesta área. Assim, pudemos aprofundar a temática da visualização de dados no jornalismo de dados, uma vez que o estudo de caso possibilita uma maneira de se investigar um tópico empírico seguindo-se de um conjunto de procedimentos pré-especificados (YIN, 2001, p. 35). Visamos nos debruçar sobre os fenômenos contemporâneos da visualização de dados e questionar a prática no jornalismo de dados no cenário nacional. O estudo de caso é sustentado por teoria e visa a reunião do maior número possível de informações por meio de diferentes técnicas de levantamento de dados, que para solucionar o problema,

ressaltam o “como” e o “por que” do fenômeno. A triangulação de informações, dados e evidências garantem a confiabilidade e a validade dos achados do estudo.

Busca-se, criativamente, apreender a totalidade de uma situação – identificar e analisar a multiplicidade de dimensões que envolvem o caso – e, de maneira engenhosa, descrever, compreender, discutir e analisar a complexidade de um caso concreto (...)” (MARTINS, 2008, p. 10).

Esta metodologia é caracterizada pelo estudo profundo e exaustivo de um ou de poucos objetos, que permitem um conhecimento mais amplo e detalhado do objeto (GIL, 1994). Exatamente o que precisamos para a compreensão aqui proposta, o conhecimento de uma realidade não conseguida plenamente por uma coleta amostral e avaliações quantitativas. De acordo com o pesquisador Robert Yin (2011) bons estudos de caso são difíceis de serem realizados. Para ele, uma pesquisa empírica fundamentada no método de estudo de caso deve ter como componentes questões do estudo, proposições, unidades de análise, lógica que une os dados às preposições e critérios para interpretar os resultados. Um estudo de caso é uma investigação empírica que estuda fenômenos contemporâneos em seu contexto real, quando: (i) não estão claramente definidos os limites entre o fenômeno e o contexto; (ii) existem mais variáveis de interesse do que pontos de dados; (iii) se baseia em várias fontes de evidências; e (iv) há proposições teóricas para realização da coleta e análise dos dados.

Yin define que existem quatro tipos de projetos na estratégia de um estudo de caso: projetos de caso único e projetos de caso múltiplos, ambos podem ser holísticos ou incorporados. Em um estudo de caso único pode se dar atenção a uma unidade de análise ou a várias unidades incorporadas. Essa distinção na quantidade de unidades é que define se um estudo de caso é holístico ou incorporado. É holístico quando se tem apenas uma unidade de análise e incorporado quando se tem mais de uma. Outro ponto a ser considerado, que distingue os dois casos, é a presença de dados quantitativos nos casos incorporados, a exemplo dos levantamentos. Dentro do método científico pode-se optar por abordagens quantitativas ou qualitativas. A pesquisa quantitativa traduz em números as opiniões e informações a serem classificadas e analisadas, utiliza-se de técnicas estatísticas. Já a pesquisa qualitativa é

descritiva, as informações não podem ser quantificáveis, os dados obtidos são analisados indutivamente e a interpretação dos fenômenos e a atribuição de significados são parte do processo de pesquisa (RODRIGUES, 2007). Neste estudo optamos pelo estudo de caso único (blog Estadão Dados), dando atenção a mais de uma unidade de análise, sendo portanto incorporado.

Optou-se pelo paradigma Jornalismo Digital de Base de Dados por ser um modelo teórico formulado para auxiliar no entendimento do papel das bases de dados no jornalismo contemporâneo que permitem a criação, a manutenção, a atualização, a disponibilização e a circulação de produtos jornalísticos digitais dinâmicos (BARBOSA, 2008 / BARBOSA & TORRES, 2013).

Para chegarmos à proposição de bases de dados como um paradigma no jornalismo digital, construímos uma abordagem teórica fundamentada num marco conceitual abrangente, indo um pouco além dos limites do jornalismo e do jornalismo digital. Buscamos embasamento nos estudos sobre os meios, sobre a internet, a nova mídia, a teoria da difusão de inovações, a teoria literária, a teoria crítica, a *web* arte, os sistemas de informação, as tecnologias da informação e comunicação, as linguagens de programação, as técnicas de descoberta de conhecimento entre os dados e os métodos para a visualização de informações. (BARBOSA, 2008, p. 219).

No estudo de caso proposto nesta dissertação observamos, em concordância com Bertocchi (2013), que a categoria visualização manifestou-se no momento de antenarração do processo de formação narrativa. Optando-se por um estudo de caso único incorporado, uma vez que foram trabalhados tanto os *projetos especiais*¹⁸¹, quanto os permanentes¹⁸² do blog Estadão Dados. No entanto, descrever um contexto ou explorar em profundidade as vivências de determinados sujeitos ainda não atendia a proposta de compreensão da visualização de dados diante das complexidades contemporâneas. Sentíamos que haviam mais elementos. **Como forma de aprofundar o estudo e tentar abarcar o fenômeno, após a qualificação realizada em 26 de junho de 2015 foram adotados princípios da Grounded Theory como método investigativo. Eles foram uma escolha apropriada para fazer emergir os processos subjacentes às afirmações dos participantes e aos fenômenos observados. Com isto, a pesquisa foi dimensionada e passou a ser uma**

¹⁸¹ Disponível em: <http://blog.estadaodados.com/projetos-estadao-dados/>

¹⁸² Disponível em: <http://blog.estadaodados.com/permanentes/>

construção criativa a partir dos dados, capaz de explicar os fenômenos pesquisados. No entanto, não adotamos esta metodologia por completo, uma vez que não é nossa intenção neste estudo elaborar uma teoria sobre a temática, mas sim estudar o objeto por outros ângulos.

Tendo ausência de hipótese prévia e de problema de pesquisa fechado, partimos da pergunta gerativa aberta, do início do estudo: O que é a visualização de dados no jornalismo? Como ela se insere na prática e na narrativa do jornalismo de dados proporcionando diferenciais relevantes? A literatura científica e os caminhos anteriores aos princípios da Grouded Theory possibilitaram um maior conhecimento sobre a área, o que ficou registrado em artigos que abordaram temas correlatos ao investigado nesta dissertação. Apesar de não ter um único caminho de investigação, há pesquisadores que defendem que em um estudo utilizando a GT não devemos partir de pressuposto algum. No entanto, é fato que um pesquisador não está totalmente livre de conceitos prévios e ideias pré-concebidas, mas ter a consciência disto tornou-se relevante para o desempenho da Grouded Theory neste estudo. Em concordância com a pesquisadora Maíra Bittencourt (2015) acreditamos que:

É impossível que um pesquisador que não seja iniciante consiga entrar em campo sem pré-noções. Se ao contrário, reconhecer essa carga de percepções pode influenciar de forma positiva. É também o que reafirma Dey (1999): “Há uma diferença entre uma mente aberta e uma cabeça vazia”. (BITTENCOURT, 2015, p. 7).

Com o olhar da Grouded Theory a pesquisa no blog Estadão Dados foi iniciada, mas antes de detalhá-la, são importantes algumas especificações da GT como metodologia e método de pesquisa.

Grounded Theory ou Teoria Fundamentada

Características

A Grounded Theory (GT) é uma metodologia, um modo de pensar, de construir a realidade social, ao mesmo tempo, em que é um método com um conjunto de procedimentos e instrumentos para trabalhar os dados da pesquisa empírica, capaz de

gerar de forma sistematizada uma teoria fundada nos dados¹⁸³ (TAROZZI, 2011; GLASSER & STRAUSS, 1967; COHEN et al, 2007). As inquietações, na maioria das vezes, brotam do campo, da prática e das observações no ambiente digital. “O êxito de uma pesquisa conduzida com a GT é uma teoria, uma interpretação racional, densa, articulada e sistemática, capaz de dar conta da realidade estudada” (TAROZZI, 2011, p. 19), com uma sólida base empírica. Alguns traços metodológicos são fundamentais para caracterizar as pesquisas na GT, são eles (Ibidem, 2011, p. 19-27):

- a) Exploração de um processo – a GT busca a exploração não de fenômenos, mas das dinâmicas percebidas em contextos. Visa a emersão dos processos sociais e psicológicos de base que subjazem aos fenômenos indagados;
- b) Amostragem teórica – na GT a amostra não se forma a priori, mas no decorrer da pesquisa, buscando as lacunas da teoria emergente;
- c) Simultaneidade entre recolher e analisar dados – na GT o momento de recolher e de analisar os dados coincide. Ela exige um acompanhamento constante da reflexão analítica com periódicos retornos ao campo;
- d) Constante comparação – a GT leva constantemente perguntas aos dados, nos vários níveis de análise. Essas perguntas buscam nexos entre dados e conceitos favorecendo o progresso da compreensão conceitual dos fenômenos estudados;
- e) Codificação a partir dos dados – a GT tem um forte enraizamento da conceituação com dados empíricos, não sendo fruto de nenhuma especulação filosófica;
- f) Elaboração de conceitos – a GT visa a elaboração de conceitos e não somente a descrição dos fatos;
- g) Produção de memorandos e de diagramas – na GT são feitos memorandos, onde é escrita a história teórica do processo analítico; onde é traçado o mapa do percurso desenvolvido e identificadas as direções a assumir no futuro.

Codificação e análise

A GT se concentra em três momentos progressivos de codificação: categorias emersas, reflexões desencadeadas e elaboração de teoria coerente. Como já

¹⁸³ Lembrando que, especificamente neste estudo, não caberá a geração de uma teoria. Nos atentamos apenas aos pressupostos da GT.

mencionado, neste estudo vamos nos restringir as duas primeiras etapas, uma vez que a proposta desta dissertação (um estudo iniciante na área) não é a elaboração de uma teoria. Retomando a GT, a codificação inicial explora analiticamente os dados, abrindo-os a todas as direções de sentido possíveis. As reflexões desencadeadas analisam elementos conceituais comuns subjacentes a porções esboçando as categorias e, posteriormente, reunindo-as em macrocategorias. A terceira é o momento de construção da teoria. Quando as categorias estão maduras, evidenciam-se os nexos que as interligam e, sobretudo, são integradas dentro de uma teoria coerente e unitária. “(...) a codificação, em todas as fases, é um processo de imersão, e o momento de abandonar o plano descritivo para entrar naquele conceitual (...)”. (TAROZZI, 2011, P. 135).

Deve-se a partir da imersão nos dados identificar macrocategorias, interligar categorias entre si e definir suas propriedades. A codificação *in vivo* visa utilizar as mesmas palavras dos participantes para dar nome aos conceitos. A codificação teórica visa uma *core category* (conceito-chave da GT), é uma categoria central que organiza o conjunto das categorias. “(...) encontrar e aprofundar (ou as) a *core category* (ies) é o objetivo da codificação teórica, a fase de codificação que se desenvolve no nível máximo de abstração conceitual” (TAROZZI, 2011, P. 140). Deve-se:

(...) analisar as relações entre as categorias e conceituar a tipologia de relação que as une; desenvolver (dedutivamente e não indutivamente) as categorias com as subcategorias com base nas propriedades e nas dimensões de cada uma; colocar as categorias em relação hierárquica, identificando macrocategorias que compreendem categorias menores. (TAROZZI, 2011, P. 143).

Escrever uma GT é um processo de teorização. Deve partir de uma revisão dos memorandos, dos diagramas que integram a teoria, e da revisão bibliográfica, a comparação entre os estudos presentes na literatura. As quatro características fundamentais da GT podem ser base para a avaliação do trabalho: ser adequada, relevante, funcionar e ser modificável.

Avaliando uma Grounded Theory

Cada pesquisa que aborda a Grounded Theory apresenta suas peculiaridades, mas existem elementos que podem ser levados em consideração na hora de avaliar um

estudo. São eles: amostragem teórica; percurso do processo; integração da teoria; profundidade; relevância; utilidade.

- Amostragem teórica: como foi selecionada a amostra? Como procedeu a amostragem teórica? Os dados são suficientes para sustentar as inferências produzidas?

- Percurso do processo: as categorias foram geradas empiricamente? Pode-se percorrer o processo de geração das categorias? É possível voltar aos principais eventos que indicaram as categorias? Existem laços sólidos (lógicos) entre a análise conduzida e os dados coletados? As evidências empíricas são suficientes para consentir ao leitor compartilhar a análise?

- Integração da teoria: os conceitos são sistematicamente interligados? As categorias são adequadamente desenvolvidas? Como e por que foi selecionada a categoria fundamental? Com base em quais razões?

- Profundidade: a teoria gerada dá conta do fenômeno estudado em sua totalidade? Foram revelados exclusivamente aspectos marginais e significados óbvios? Os resultados são ligados ao debate científico atual sobre o tema?

- Relevância: as categorias e o processo apresentado abrem novas perspectivas? São significativos? Desafiam o saber consolidado sobre o tema? São conceitualmente densos? Tem um forte poder explicativo? A teoria e os conceitos apresentados duram no tempo? Abrem ulteriores pesquisas?

- Utilidade: a teoria tem sentido para os participantes? Oferece interpretações aplicáveis pelos operadores? Intercepta questões relevantes no debate profissional? Surge pistas para mudanças e inovação?

De posse desses pressupostos da Grounded Theory partimos para a explanação de como se deu a aplicação no nosso objeto de estudo.

Aplicando a Grounded Theory no blog Estadão Dados

Três instrumentos foram de suma importância no fazer desta pesquisa e que na GT assumem características particulares:

- Observação netnográfica – focalizamos imediatamente na observação dos fenômenos e, sobretudo, dos elementos de processo definidos na pergunta de pesquisa, dando em um primeiro momento, menor peso à descrição do contexto;
- Entrevista – depois da entrevista escrevemos rapidamente notas de campo. E neste estudo, optou-se por não gravá-las já que foram feitos acompanhamentos mais longos dentro da redação do jornal no Blog Estadão Dados;
- Documentos e análises de texto – foram úteis para o confronto com os dados de campo. Já havíamos feito fichamentos de dezenas estudos ao longo dos primeiros 15 meses de pesquisa.

Acompanhamento do blog

Apesar de observar desde 2013 o Blog, o acompanhamento mais sistemático aconteceu de janeiro a dezembro de 2015.

Entrevistas semi-estruturadas

Para a melhor compreensão do cenário da prática de visualização de dados foram feitas entrevistas em profundidade com jornalistas e demais profissionais indicados pelo jornal O Estado de S. Paulo. O objetivo foi captar qualitativamente as múltiplas vozes dos envolvidos no cotidiano desta prática. Para Jensen e Jankowski, “os enfoques qualitativos examinam a produção de significado como um processo, que se contextualiza e se integra com as mais amplas práticas sociais e culturais” (JENSEN & JANKOWSKI, 1993, p. 13). É tanto uma ferramenta quanto um objeto de análise.

Ao comunicar-se por meio da linguagem, o entrevistador e o entrevistado negociam um certo conhecimento da matéria em questão que, mais tarde,

em forma de transcrições, se converte em objeto de análise lingüística e de interpretação textual (JENSEN & JANKOWSKI, 1993, p. 44).

De acordo com o pesquisador Michel Thiollent (1980) uma estratégia metodológica não nos possibilita abarcar toda a realidade. As verdades são construídas e parciais e estão relacionadas com os aspectos teórico-metodológicos de uma pesquisa, ou seja, é falsa a neutralidade das enquetes sociológicas. A utilização de uma técnica de pesquisa envolve limitações – aspectos favoráveis e desfavoráveis. A opção de realizar a entrevista em profundidade se deu pela facilidade de entrar em contato e entrevistar as pessoas envolvidas no cotidiano da redação e por permitir uma relação de confiança entre pesquisador-pesquisado. As entrevistas foram realizadas a partir de questões semi-estruturadas que tiveram por objetivo saber:

- 1) formação profissional: qual a formação dos envolvidos nos projetos em visualização de dados; como chegou a trabalhar com dados e jornalismo; com quais ferramentas de visualização trabalha e como aprendeu a usá-las;
- 2) construção noticiosa: como geralmente acontece a captura dos dados; quais as fontes utilizadas; como os dados são agrupados; quais as ferramentas utilizadas no veículo em que trabalha e como se chega aos assuntos abordados;
- 3) investimentos: quais os investimentos da instituição na área em recursos tecnológicos, contratação de mão-de-obra e qualificação profissional;
- 4) dificuldade em lidar com dados: quais as principais dificuldades.

Caminhos de estruturação

No próximo capítulo será apresentada a estruturação já contextualizada aos dados. Será o momento de construção, revisão, desconstrução, retorno ao campo, busca de lacunas descobertas. Até encontrar a *core category*, a categoria central onde todas as outras categorias devem estar ligadas. De acordo com o sociólogo americano Barney Glaser, um dos fundadores da Grounded Theory, é preciso integrar as categorias

dando atenção para as causa, os contexto, a contingência, as consequências, as covariáveis e as condições. “Esse modelo de estrutura era defendido com bastante rigidez pelo autor, quando de sua publicação, mas o entendimento atual é que são pistas que darão um bom sustento para esta, sendo possível trabalhar também com outras questões” (BITTENCOURT, 2015, p. 11). Entre os pontos que visamos aprofundar estão:

- a) Delimitações e categorias emersas: quais peças jornalísticas serão analisadas, por quais aspectos e porquê. Tendo em mente a percepção macro (do Blog como um todo¹⁸⁴) e micro (no que diz respeito as peculiaridades da prática) no ambiente redacional. Primeiramente o Blog Estadão Dados, posteriormente os exemplos das peças jornalísticas Basômetro e Farra do Fies; o cenário em que nosso objeto de estudo está inserido; quais as influências internas e externas do fazer;
- b) Reflexões desencadeadas: pontos jornalísticos destacados e desafios do novo;
- c) Sustentação teórica vinculada a *core category*: diálogo entre as teorias emergentes destacadas na pesquisa e o objeto de estudo.

Os dados coletados neste estudo são frutos de visitas a redação do jornal o Estado de S. Paulo, especificamente ao blog Estadão Dados e ao Acervo (setor que responde pelo armazenamento de conteúdos do jornal), onde optamos por não gravar as conversas, mas sim nos aproximarmos das fontes da pesquisa, como apontado na página 140. Além das conversas presencialmente, foram tiradas dúvidas por email e telefone com os coordenadores dos dois setores, respectivamente, os jornalistas José Roberto de Toledo e Edmundo Leite. Os esclarecimentos presenciais foram feitos também pelos jornalistas Daniel Bramatti e Rodrigo Burgarelli e pelo cientista de dados Guilherme Duarte. Complementaram os estudos anotações e bate-papos, em mini-cursos e congressos, com jornalistas que fizeram parte do Blog ou que compreendem a execução de etapas como mineração, extração, conversão de dados e

¹⁸⁴ Não vamos abordar a história do jornal do qual o blog faz parte, pois demandaria tempo e não é o objeto do nosso estudo. Para isto há farta referência bibliográfica uma vez que falamos de um veículo tradicional e o mais antigo em circulação na cidade de São Paulo. Informações complementares encontram-se disponíveis no site do veículo em: <http://www.estadao.com.br/historico/resumo/contil.htm>. Acesso em 23 de fevereiro de 2016.

acessos a bases de dados públicas. Os nomes e eventos são citados ao longo do próximo capítulo em notas de rodapé.

Considerações

Este quarto capítulo sintetizou as dificuldades dos estudos em Comunicação na contemporaneidade e o olhar que damos aos fenômenos neste contexto. Um olhar que tem a comunicação como centro, que perpassa todas as atividades comunicativas e informativas de forma transversal, sejam elas analógicas ou não. Um olhar que requer resiliências diante das abordagens teórico-metodológicas para que as mesmas consigam abarcar e sustentar as reflexões referentes ao dinâmico campo de estudos. Elementos que justificam nossas escolhas metodológicas que envolve cruzamentos de distintos caminhos informativos: pesquisa bibliográfica, observação não-participante, estudo de caso, fundamentos da Grounded Theory, cursos, disciplinas, grupo de pesquisa Com+, palestras, estágio de um ano na disciplina Jornalismo on-line, momentos que foram descritos anteriormente.

Temos por meta compreender a formação dos profissionais envolvidos (como chegou a trabalhar com dados e jornalismo; com quais ferramentas de visualização trabalha e como aprendeu a usá-las); como se dão as construções noticiosas (captura dos dados, fontes utilizadas, como os dados são agrupados, quais as ferramentas utilizadas); os investimentos na área (em recursos tecnológicos, contratação de mão-de-obra, qualificação profissional) e saber das dificuldades cotidianas em lidar com dados. Para a partir destas considerações aprofundar a discussão na busca: de delimitações e categorias emergentes (percepções macro e micro no ambiente redacional); de reflexões desencadeadas (pontos jornalísticos destacados e desafios do novo) e da elaboração de uma sustentação teórica vinculada a *core category*, com o diálogo entre as teorias emergentes destacadas na pesquisa e no objeto de estudo.

O capítulo V, portanto, faz uma inversão, onde diante dos dados as teorias são questionadas. O capítulo tem por objetivo exemplificar a prática da visualização de dados a partir dos conteúdos trabalhados no blog Estadão Dados e a meta de agrupar elementos que ajudem a responder a pergunta problema desta pesquisa: O que é a

visualização de dados no jornalismo? Como ela se insere na prática e na narrativa do jornalismo de dados proporcionando diferenciais relevantes?

V. BLOG ESTADÃO DADOS

Resumo: Este capítulo tem por objetivo exemplificar a prática da visualização de dados a partir dos conteúdos trabalhados no blog Estadão Dados. Parte-se de uma descrição geral do blog e peculiaridades de cada seção, para em seguida abordar os processos incluídos nas peças jornalísticas.

Palavras-chave: Blog Estadão Dados; ferramenta de visualização; Jornalismo de Dados

Lindando com os dados

Houve um tempo em que, para se saber o voto de cada deputado em relação a um determinado projeto, era preciso comparecer a uma salinha localizada abaixo do plenário da Câmara, entrar em uma fila e solicitar ao atendente um relatório emitido pelo painel eletrônico. Lá funcionava uma única impressora matricial, cuja “cabeça” golpeava com agulhas uma fita impregnada de tinta, marcando no papel pontos em linhas horizontais que, após alguns milhares de vaivéns, desenhavam 513 nomes, siglas partidárias e, ao lado, os registros de sim, não abstenção e ausente (BRAMATTI, 2014, p. 219).

De acordo com o relato do jornalista Daniel Bramatti (acima) é perceptível a dificuldade de correlacionar as informações e se fazer uma análise do comportamento dos parlamentares e das bancadas. O depoimento não é de um passado distante, mas dos primeiros anos do governo Fernando Henrique Cardoso (que esteve a frente da presidência do país entre 1º de janeiro de 1995 e 1º de janeiro de 2003), onde houveram alterações da Constituição como a quebra do monopólio estatal do petróleo e da privatização das telecomunicações. Com a digitalização, estes dados passaram a ser disponíveis em arquivos de texto ou de banco de dados, mas “como a composição dos presentes variava a cada votação, nenhuma lista era igual à outra, o que, mais uma vez, praticamente inviabilizava a análise de séries históricas” (BRAMATTI, 2014, p. 220). O panorama só começou a mudar de forma significativa uma década

depois, quando o movimento pela publicação de dados abertos deu seus primeiros passos no Congresso.

Hoje, com a comunicação em rede e o manuseio de grandes volumes de dados, as informações são passíveis de serem agrupadas e conversáveis entre si com maior facilidade. Isto não quer dizer que seja uma tarefa fácil. Requer habilidades, mas é possível. Como exemplo, há a construção de ferramentas de visualização como o Basômetro¹⁸⁵ (citado brevemente no capítulo II), que estreou em maio de 2012, juntamente com o blog Estadão Dados, do Jornal O Estado de S. Paulo. Foi a primeira ferramenta criada pelo grupo, originária do primeiro núcleo de jornalismo de uma consolidada redação no Brasil. Nasceu sob a coordenação do jornalista, especialista em política, José Roberto de Toledo, que atualmente também é diretor da Abraji.

Os passos deste estudo nos levaram por dois caminhos. Em um primeiro momento: uma abordagem macro onde buscamos compreender as dinâmicas do trabalho da visualização de dados dentro da redação. Uma atividade que possui um ciclo de produção e consumo próprios. Mas, como o objetivo desta pesquisa era ir além, encontrar outras nuances que remetem ao que é a visualização de dados no jornalismo, achamos importante destacar o trabalho desenvolvido pelos jornalistas do Blog a partir de um outro prisma, aproximando o olhar. O segundo caminho, portanto, foi aprofundar as discussões de forma mais sistemática a partir de dois exemplos:

- do próprio Basômetro por ser uma ferramenta em visualização disponibilizada para o acesso do público e pela sua importância por coincidir com a estreia do blog Estadão Dados;
- e da série de reportagens Farra do Fies¹⁸⁶, por envolver a formação de uma base de dados própria (originária do cruzamento de outros bancos de dados) para a produção de conteúdos relacionados ao financiamento do ensino superior no país e pelos desdobramentos que esta série de reportagens obteve.

¹⁸⁵ Aprofundaremos a discussão sobre o Basômetro ao longo deste capítulo.

¹⁸⁶ <http://blog.estadaodados.com/?s=fies>

A partir deste novo afunilamento visamos encontrar elementos para nos auxiliar a sustentar as respostas desta pesquisa.

Blog Estadão Dados

O blog Estadão Dados tem um núcleo estruturado e autônomo dentro da redação do Jornal O Estado de S. Paulo. É especializado em reportagens baseadas em estatísticas e no desenvolvimento de projetos especiais de visualização de dados. No Blog são observados gráficos e animações sobre temas cotidianos, além de análises especiais feitas a partir do cruzamentos de dados. A equipe, atualmente, é formada por jornalistas e programadores, onde atuam diretamente os profissionais: José Roberto de Toledo – coordenador do Blog; Rodrigo Burgarelli - repórter e analista de dados; Guilherme Duarte – cientista de dados e Daniel Bramatti - repórter.

Figura V. 15 – Estadodados.com, *home* do dia 14 de dezembro de 2015.

Fonte: <http://blog.estadaodados.com/>

São diversas as frentes de trabalho desenvolvidas. No Blog encontram-se as seções Gráfico do Dia, Permanentes, Séries, Eleições, 96xSP e Projetos Especiais. Além dos espaços para comentários, compartilhamento via redes sociais e links para posts similares, que levam para as diversas editorias que repercutiram os conteúdos publicados. As seções se destinam a:

- Gráfico do Dia – trazer notícias cotidianas, com diversas temáticas, mas principalmente políticas;
- Permanentes – incluir trabalhos atemporais, estatísticas que são constantemente atualizadas;
- Séries – diversos conteúdos que focam em um tema específico, são mais aprofundados e com usos de recursos mais complexos;
- Eleições – publicações referentes a cobertura das eleições de 2014;
- 96xSP – matérias que abordam as características dos 96 distritos paulistanos, segundo o Censo 2010;
- Projetos Especiais – trabalhos mais complexos que utilizam, muitas vezes, ferramentas de visualização.

A produção não tem a obrigatoriedade de ser diária, mesmo na seção denominada Gráfico do Dia. Em coleta feita em setembro de 2015, por exemplo, somente esta seção estava com atualização no mesmo ano. Isso não quer dizer que o conteúdo seja defasado, apenas que há outra lógica que rege a produção específica destas peças jornalísticas que não a produção diária, como estamos acostumados a ver nos veículos tradicionais de mídia (retomaremos este assunto mais a frente). Em coleta feita no dia 03 de setembro de 2015:

- Gráfico do Dia – haviam 9 páginas, 8 delas com 20 gráficos e uma com 3. Total de 163 trabalhos;
- Permanentes – haviam duas páginas, uma 1 com 20 e outra com 4 matérias. Total de 24 peças;
- Séries – haviam 3 páginas, 2 com 20 e uma com 8 matérias. Total de 48 peças;
- Eleições – haviam 8 matérias;
- 96xSP – eram 2 páginas, uma com 20 e outra com 5 matérias. Total de 25 peças;

- Projetos Especiais – encontravam-se 12 projetos.

Em nova observação feita no Blog em 2016, apenas a seção Gráficos do Dia teve alteração no número de matérias, subindo de 163, em setembro de 2015, para 168 em 18 de fevereiro de 2016. As demais encontravam-se sem novas publicações. Diante desta breve apresentação do Blog, vamos a sua história que coincide com a criação da ferramenta Basômetro.

Basômetro

O Basômetro analisa a taxa de governismo¹⁸⁷ dos políticos da Câmara Federal e do Senado através dos dados das votações realizadas nas casas, podendo consultar a proposição dos votos desde 2003. “Não é um infográfico, não é um banco de dados, não é uma tabela, mas é tudo isso ao mesmo tempo” (BOUNEGRU et.al, 2012, p. 102), definiu o livro *Data Journalism Handbook*, na versão em português intitulado *Manual de Jornalismo de Dados*. Com uma interface intuitiva, o Basômetro permite a aplicação de filtros e a seleção de períodos de tempo ou votações específicas, fazendo assim com que cada um crie um caminho narrativo próprio. Possibilita ao usuário visualizar quão a favor ou contra certo deputado, senador ou legenda é do governo.

O Basômetro organizou centenas de milhares de dados que eram mero ruído dos sites do Congresso em uma interface gráfica que os cientistas políticos puderam e podem usar para fazer suas análises. (TOLEDO, 2014, p. 9)

Na página da ferramenta Basômetro há um tutorial de uso para auxiliar na leitura dos dados, onde podem ser feitas as escolhas de que votações acompanhar, por exemplo: no Senado ou na Câmara dos Deputados. Trazendo o recorte das votações na câmara, as cores das bolinhas indicam os partidos de cada deputado. Na legenda ao lado esquerdo aparecem quais as cores de cada partido. Ela funciona também como filtro, quando clicada mostra o comportamento do partido nas votações. Cada círculo dentro

¹⁸⁷ A taxa de governismo é a medição do apoio ou não dos deputados ao governo. Por exemplo, a taxa de governismo será maior quanto maior for o número de votos pró-governo em relação ao total de votos do período.

do gráfico representa um deputado. Clicando ou tocando nestes círculos descobre-se as votações do candidato. Sendo possível ver quantos por cento cada um votou a favor do governo. Quanto mais pra cima do gráfico estiverem as bolinhas, mais governista foi o candidato. Quanto para baixo, mais oposicionista. É possível restringir a análise em um período específico de tempo e fazer inúmeras outras explorações a partir dos dados disponíveis. A evolução do apoio ao governo ao longo do tempo pode ser feita acionando a tecla *play*, no campo inferior direito. Ela mostra a mudança do comportamento dos deputados e dos partidos ao longo das votações. Na página é possível acompanhar análises em vídeo já feitas pela equipe, como a do núcleo duro de Dilma, daqueles que votaram de acordo com a orientação do líder do governo em 90% dos casos ou mais vezes.

Figura V. 16 – Usabilidade da ferramenta Basômetro – taxa de governismo de candidato do PSC/PB

Fonte: estadaodados.com/basometro, 2016

Agora, como começou o Basômetro? Em 2011, a Câmara dos Deputados e do Senado divulgou o projeto Dados Abertos¹⁸⁸ onde passou a disponibilizar digitalmente dados

¹⁸⁸ <http://dados.gov.br/>

legislativos permitindo conhecimento, apropriação e reutilização das informações por outros órgãos públicos, entidades da sociedade civil e demais interessados. E como aponta Daniel Bramatti, “não demorou para que cidadãos começassem a manifestar sua criatividade. Em janeiro de 2012, Leonardo Leite, mestrando em Ciências da Computação e integrante do PoliGNU, da USP, lançou o projeto “Camaraws””¹⁸⁹ (BRAMATTI, 2014, p. 223). A ideia de criar o Basômetro surgiu exatamente a partir de um post, no *Facebook*, que falava sobre este trabalho desenvolvido no laboratório da Escola Politécnica, da Universidade de São Paulo. Era uma ferramenta de visualização que trazia dados de votação, que mostrava a direita e a esquerda governista do país.

Graças ao novo formato de publicação do projeto Dados Abertos, Leonardo Leite pôde agregar dados de várias votações e consolidá-los segundo diferentes critérios – por partido, por estado, etc. Em seguida comparou o comportamento dos partidos, dois a dois, para revelar o índice de semelhança entre eles. Quando tomou conhecimento sobre o *Camaraws*, o jornalista político José Roberto de Toledo entrou em contato com a equipe da PoliGNU para fazer algo em conjunto. Seria um bom desafio para iniciar as atividades do blog Estadão Dados que já estava em formação.

Os desdobramentos do fazer

No caso do Basômetro, primeiramente foi preciso conhecimento em programação para a análise de grandes quantidades de dados que não são suportados por, exemplo, em programas como *Excel* que possui um limite do que consegue analisar. A Câmara e o Senado, diante da política de dados abertos, já disponibilizava via API's os votos dos parlamentares na linguagem *Python*. O Basômetro focou nesta mesma linguagem (que é de alto nível) para a raspagem dos dados e em *JavaScript* (que roda em códigos HTML) para gerar os gráficos e manter a interatividade da ferramenta assumindo o papel humanamente impossível de analisar todos os votos e chegar a resultados passíveis de compreensão.

¹⁸⁹ O “ws” do nome *Camaraws* vem de “web services”. O projeto foi rebatizado, posteriormente, para Radar Parlamentar e está disponível no endereço: <http://radarparlamentar.polignu.org/>. Acesso em 6 de fevereiro de 2016.

A programação esteve presente em todo o processo informacional, na apuração, na produção e na apresentação dos dados. No entanto, os profissionais envolvidos foram os definidores das variáveis destacadas para a análise de cada situação. “Foi um trabalho intenso, que passou por diversas releituras, inclusive da fuga da militância política. Foram adotadas as nomenclaturas governo e oposição. Foram alimentados os votos de deputado a deputado e de senador a senador”, disse Toledo¹⁹⁰.

Foram consideradas todas as votações nominais, as únicas em que o voto individual é computado, que ocorreram desde o início de 2011. No caso da Câmara, não entraram no levantamento os casos em que o governo não orientou os deputados sobre como votar. No Senado, foi usado como referência o voto do líder do governo na Casa (ESTADÃO DADOS, 2012)¹⁹¹.

Na sua estreia, o *software* da ferramenta foi adaptado para absorver dados sobre 825 votações e cerca de 420 mil votos na Câmara nos dois mandatos do presidente Luiz Inácio Lula da Silva. “Só é preciso fazer uma vez o código, depois é só repetir ou alterar as variáveis. Foram discutidas as bases, em escalas de 0 a 100 e definido que o código seria aberto”, complementou Toledo.

Os jornalistas Bruno Lupion e Amanda Rossi organizaram, posteriormente, uma base de dados com todas as votações da Câmara e do Senado no governo Dilma Rousseff com a posição de cada parlamentar em relação à orientação do líder do governo. Paralelamente a isso, o programador e designer Carlos Lemos construiu um *software* para processar estes dados. Como o objetivo era permitir que os usuários fizessem diversas correlações e leituras dos dados, a arquitetura, desde o princípio, levou em consideração a interatividade e uma interface que possibilitasse ao usuário analisar o comportamento de indivíduos, partidos e bancadas estaduais em votações isoladas ou em qualquer período do governo Dilma. Com isto, o Basômetro possibilitou inúmeras informações até então inéditas.

(...) Reportagens revelaram que o “núcleo duro” de Dilma – formado por parlamentares que votavam com o governo em 90% das vezes ou mais – havia encolhido no segundo ano do mandato, e que o PMDB era o principal responsável pela dispersão da base aliada. (...) O Basômetro também permitiu ver quais eram os deputados e senadores

¹⁹⁰ Em visita a redação de O Estado em 17 de novembro de 2015.

¹⁹¹ Reportagem de O Estado de S. Paulo, intitulada “Ferramenta mede taxa de governismo de deputados”. Publicada em 13 de maio de 2012. Disponível em: <http://politica.estadao.com.br/noticias/eleicoes,ferramenta-mede-taxa-de-governismo-de-deputados-imp-,872318>. Acesso em 6 de dezembro de 2015.

“independentes”, que assumiam posições divergentes das de seus partidos. (BRAMATTI, 2014, p. 225).

O Basômetro encontra-se na seção Permanentes do blog Estadão Dados e foi o tema do livro *Análise política & jornalismo de dados*, sob a organização do doutor em ciência política pela USP, Humberto Dantas, pelo jornalista e coordenador do blog Estadão Dados José Roberto de Toledo e pelo doutor em ciências sociais da PUC-SP Marco Antônio Carvalho Teixeira. O livro, de 2014, reuniu cientistas políticos e jornalistas para discutir hipóteses do cenário da política nacional. Após o conhecimento sobre este projeto, vamos a outro exemplo, que não resultou em uma ferramenta, mas onde foi consolidado um banco de dados para a construção de uma série de reportagens-denúncia.

Farra do Fies

Publicada em fevereiro de 2015, a reportagem Farra do Fies¹⁹² investigou os gastos do governo federal com o Financiamento Estudantil (FIES). Ela mostrou que o investimento federal no programa aumentou 13 vezes entre 2010 e 2014. No entanto, também apontou que caiu o ritmo do crescimento de matrículas nas faculdades particulares. A média anual passou de 5% entre 2003 e 2009, para 3% de 2010 até 2013. O trabalho mostrou ainda que subiram as transferências de recursos da União para grupos de educação após alterações nas regras do Fies. Em 2010, não havia nenhuma empresa de educação entre as 70 que mais recebiam do governo federal. Já em 2014, sete constavam na lista.

Paulo Saldaña, Rodrigo Burgarelli e José Roberto de Toledo conquistaram a categoria principal da 60ª edição do ExxonMobil 2015 (antigo Prêmio Esso de Jornalismo) com reportagem a Farra do Fies. A comissão responsável pelo prêmio destacou o fato de o especial trazer, além de entrevistas e análises de documentos, o cruzamento de grandes bancos de dados, como o do Censo de Educação Superior e o do Portal de Transparência, além de registros de mais de 300 instituições de ensino do país. “O esforço revelou, entre outros aspectos até então ocultos, que o programa consumiu R\$

¹⁹² <http://blog.estadaodados.com/?s=fies>

28 bilhões em quatro anos, endividando alunos que dificilmente terão condições de ressarcir os cofres públicos”, apontou o anúncio da premiação¹⁹³.

O Fies foi criado em 1999 para emprestar dinheiro para alunos cursarem faculdades particulares. Com o intuito de aumentar o número de alunos, em 2010 houveram mudanças nos contratos onde os juros caíram de 6,5% para 3,4% ao ano, valor este abaixo da inflação na época. Além disso, o financiamento pôde ser obtido a qualquer momento, foi relaxada a exigência de fiador e o prazo de quitação foi prolongado. Com estas medidas o Fies teve uma explosão de contratos. A principal consequência das novas regras foi o crescimento nas transferências de alunos particulares para alunos Fies nos grupos de educação.

O Kroton-Anhanguera, por exemplo, foi a empresa que mais recebeu pagamentos do governo federal em 2014. Doze mantenedoras do grupo receberam juntas mais de R\$ 2 bilhões - o dobro do que a Embraer, que fabrica aviões militares, e a Odebrecht, responsável por dezenas de obras pelo país (TOLEDO et al. 2015)¹⁹⁴.

O número de alunos no Fies subiu 448%, ou seja, de 150 mil, em 2010, para 827 mil em 2013, último ano em que há dados do Ministério da Educação. No entanto, o total de alunos na rede privada subiu apenas 13% - de 3,9 milhões para 4,4 milhões. “A explicação é que muitas faculdades passaram a incentivar alunos já matriculados a não pagar a própria mensalidade, mas a entrar no Fies – que, por sua vez, repassa os valores diretamente para as instituições, sem atraso ou inadimplência” (TOLEDO et al. 2015).

Sete infográficos interativos foram desenvolvidos para explicar a farra do financiamento estudantil e encontram-se disponíveis para consulta no blog Estadão Dados. Na sequência, vamos aos desdobramentos de como se chegou a série de reportagens Farra do Fies.

¹⁹³ Reportagem “Pelo segundo ano consecutivo, ‘Estado’ vence principal premio de jornalismo do país”, publicada pelo jornal O Estado de S. Paulo em 19 de outubro de 2015. Disponível em: <http://politica.estadao.com.br/noticias/geral,pelo-segundo-ano-consecutivo-estado-vence-o-principal-premio-de-jornalismo-no-brasil,1782259>. Acesso em 20 de dezembro de 2015.

¹⁹⁴ Reportagem “Gasto com o Fies cresce 13 vezes e chega a R\$ 13,4 bi, mas ritmo de matrículas cai”, publicada pelo jornal O Estado de S. Paulo, em 15 de fevereiro de 2015. Disponível em: <http://sao-paulo.estadao.com.br/noticias/geral,gasto-com-fies-cresce-13-vezes-e-chega-a-r-13-4-bi-mas-ritmo-de-matriculas-cai-imp-,1634714>. Acesso em 20 de dezembro de 2015.

Figura V. 17 – Série de infográficos interativos explicam financiamento estudantil

Gastos com o Fies explodem

Os grupos educacionais lucram como nunca

Mas o número de alunos não aumenta no mesmo ritmo

Na verdade, o ritmo de crescimento do ensino superior privado caiu

Isso acontece, em parte, porque pegar o Fies é lucrativo

Alguns grupos dependem mais do Fies que outros

E, na média, as mensalidades estão encarecendo

Fonte: <http://blog.estadaodados.com/?s=FIES>, 2015

Os desdobramentos do fazer

De acordo com José Roberto de Toledo, a pauta sobre o Fies surgiu ao acaso, quando ele observava os dados do Portal de Transparência¹⁹⁵, do Governo Federal, e achou curioso o fato de entre as empresas que receberam mais dinheiro público se encontrarem instituições de ensino privadas. Por que isto? Por que estas instituições em especial? Por que estes valores? O faro jornalístico indicou que poderia haver uma pauta naquela primeira informação. A partir daí a indagação foi encaminhada para o jornalista Rodrigo Burgarelli que foi levantar os valores: quando a instituição começou a receber e qual a trajetória de recebimento ao longo dos anos. Iniciou-se o processo de pergunta aos dados. Por que, a partir de 2010, há um enorme crescimento nestes valores? O que aconteceu? Neste momento, o jornalista Paulo Saldaña, editor de Educação do jornal O Estado de S. Paulo, trouxe os esclarecimentos que faltavam sobre a mudança das regras. Agora era seguir pois havia sim uma pauta. A equipe do

¹⁹⁵ <http://www.portaltransparencia.gov.br/>

Estadão Dados, especialmente o Rodrigo Burgarelli, precisou primeiramente escrever em código, para daí escrever a matéria na língua portuguesa.

Entrevistamos milhões de registros do Orçamento Geral da União e do Censo da Educação Superior. Aos dados se juntaram as histórias de alunos e escolas, tão bem apuradas e contadas por Paulo Saldaña. Sem a junção da técnica com o conhecimento, sem combinar a análise estatística com a experiência profunda em educação, a Farra do Fies não existiria, afirmou. (TOLEDO apud THOMÉ, 2015)¹⁹⁶

Em entrevista no dia 1º de março de 2016, Toledo comentou sobre os procedimentos adotados no desdobramento da reportagem, sobre como foi elaborada a base de dados e suas correlações. Em suas palavras foi perceptível a presença de cinco momentos na construção do sistema narrativo (TOLEDO, 2016):

1º) “Não uma só, mas apareceram várias empresas educacionais. O que era uma anomalia. Quase a totalidade dos cursos era do Fies. A partir daí capturamos os dados do Portal de Transparência. Não sei dizer se foi exatamente uma lista ou qual formato. Conseguimos alguns dados e começamos a trabalhar num banco de dados em *MySQL*¹⁹⁷. Fizemos a correlação dos dados, quem tinha recebido, quanto”.

2º) “Saldaña fez uma correlação com uma outra base de dados do MEC, para ver quais nomes de empresas pertenciam a um mesmo grupo. Daí fizemos uma agregação de dados e fomos vendo quais instituições eram as maiores recebedoras. Com base nisso conseguimos fazer um ranking”.

3º) “Capturamos a base do Fies para ver quantos alunos tinham o financiamento. Calculamos a taxa de cada uma delas. Conseguimos ver o acréscimo significativo”.

4º) “Captamos uma outra base do Censo do Ensino Superior, do MEC. Fizemos uma análise para ver a evolução do número de matrículas no ensino superior pra constatar que não havia nenhum acréscimo significativo no número de alunos em decorrência do Fies. Ele estava praticamente estável”.

¹⁹⁶ Reportagem “Jornalistas do ‘Estado’ recebem no Rio prêmio Esso de jornalismo”. Publicada no dia 13 de novembro de 2015. Disponível em: <http://politica.estadao.com.br/noticias/geral,jornalistas-do-estado-recebem-no-rio-premio-esso-de-jornalismo,10000001888>. Acesso em 20 de dezembro de 2015.

¹⁹⁷ O MySQL é um sistema de gerenciamento de banco de dados (SGBD), que utiliza a linguagem SQL (Linguagem de Consulta Estruturada, do inglês *Structured Query Language*) como interface. É atualmente um dos bancos de dados mais populares, com mais de 10 milhões de instalações pelo mundo. Disponível em: <https://pt.wikipedia.org/wiki/MySQL>. Acesso em 07 de março de 2016.

5º) “Fizemos uma outra base para ver o preço médio de mensalidades pra mostrar que vinham caindo e voltou a crescer a partir do Fies. Foram junções de bases de dados. Chegamos aos resultados que explodem gastos frente ao Fies por parte do governo, mas que existem os mesmos número de alunos.

A equipe esteve envolvida na construção e correlação das bases de dados por aproximadamente dois meses e meio. As reportagens foram publicadas no decorrer de nove dias, por conta do número de informações e desdobramentos gerados. Diante de uma visão mais próxima dos processos de execução do Basômetro e da Farra do Fies, tendo os dados coletados nas conversas, entrevistas com a equipe do Blog e a investigação sobre o assunto que aconteceu ao longo da realização do mestrado, iniciamos as reflexões sobre a visualização de dados e o jornalismo de dados a partir das informações obtidas. O objetivo posterior será o de discutí-las juntamente com as referências bibliográficas adotadas na pesquisa.

Reflexões desencadeadas

O caminho para chegar ao blog Estadão Dados vem do acompanhamento desde 2013 dos conteúdos em visualização de dados em âmbito nacional e internacional. Como já mencionado no capítulo anterior, foram inúmeras buscas on-line, palestras, participação em MOOC's na área, acompanhamento de pesquisas, artigos, livros, defesas de dissertação que abordaram a temática. O estudo de caso do Blog se deu por ele ter tanto conteúdos gráficos, quanto projetos em visualização, pelas constantes atualizações de conteúdo e por estar em um cenário brasileiro de investimentos nesta área.

O primeiro contato com a equipe do Estadão Dados aconteceu a partir das oficinas em visualização de dados cursadas no 10º Congresso Internacional de Jornalismo Investigativo da Abraji, realizado em agosto de 2015, em São Paulo. As visitas à redação foram agendadas via email e logo na primeira algumas percepções surgiram: o grupo é muito pró-ativo e formado por pessoas que gostam, se envolvem e se dedicam ao que fazem. Os diversos conhecimentos são sempre bem-vindos,

indiferentemente da área, e há inúmeras buscas por troca de experiências. Como já mencionado, as rotinas produtivas são outras.

A produção não tem a obrigatoriedade de ser diária, mas sim de entregar conteúdos aprofundados, contextualizados e diferenciados na forma de contar histórias a partir de grandes bases de dados, de histórias “ocultas”. É uma equipe que tem o autodidatismo presente e a não-linearidade da escrita em trabalhos que são essencialmente desenvolvidos em grupo. É uma equipe afinada com as tendências contemporâneas e as demandas do exercício da profissão num cenário Big Data. Há uso de grandes bases de dados, busca por tutoriais e exemplos de quem já fez, tendo como elemento fundante os testes sobre os diferentes caminhos de construções narrativas.

As visitas posteriores, entrevistas e observações sistemáticas nos levaram aos pontos jornalísticos que se destacavam e aos desafios de se trabalhar com a visualização de dados. Vamos a eles!

Pontos jornalísticos destacados

Alguns pontos tornam-se perceptíveis no acompanhamento mais próximo dos trabalhos desenvolvidos no Blog. Não há uma fórmula para a execução dos trabalhos, cada peça tem suas peculiaridades. Os pontos jornalísticos que destacamos de imediato foram: finalidade do conteúdo; testes, adaptações e autodidatismo; faro jornalístico e formação da equipe; armazenamento e recuperação de dados (datalização); solução e validação das peças.

Finalidade do conteúdo

A finalidade de uso do conteúdo define o design em que o mesmo chegará ao leitor. Seja uma ferramenta, como o Basômetro, que serve para análise e consulta na base de dados ou como infográficos, estáticos ou dinâmicos, que selecionam e apresentam de forma visual o conteúdo para os leitores. São feitas escolhas na hora da apresentação dos conteúdos. Nem tudo será mostrado, afinal são grandes bases de dados que

acabam sendo formadas em cada peça. Há visualizações que funcionarão como base de dados somente para a própria equipe e uso em pautas futuras. O leque de opções de visualização em uma ferramenta como o Basômetro, por exemplo, é maior, possibilitando que o leitor faça diferentes correlações diante do banco de dados disponível e utilizando-o como fonte, seja para outras reportagens, pesquisas acadêmicas, curiosidade.

A reportagem Farra do Fies é um exemplo do resultado da construção de uma base de dados que reuniu mais de 40 mil linhas de informações de estudantes, com o cruzamento de informações do Censo de Educação Superior, do Portal de Transparência do governo federal, além de registros de mais de 300 instituições de ensino superior do país. Esta base de dados não foi integralmente disponibilizada para o público, mas serviu de fonte de consulta para a produção de reportagens do núcleo do Estadão Dados.

A matéria *Qual partido tem mais deputados com problema na Justiça?*¹⁹⁸ é um exemplo de infografia que a partir de gráficos de barras interativos mostra primeiramente quais são réus e investigados e, em um segundo momento, qual a proporção da bancada com problemas na justiça. Os dados são do Congresso em Foco¹⁹⁹ que realizou um levantamento de processos e inquéritos que correm no Supremo Tribunal Federal e envolvem deputados e senadores da atual legislatura.

O material foi disponibilizado para consulta em agosto de 2015 e no mesmo mês, usando Infogr.am²⁰⁰, parte do conteúdo foi disponibilizado pela equipe do Estadão Dados. Uma forma de apresentação de um conteúdo complexo, relevante pela interligação e possibilidade de cruzamento de dados, de forma visual e que facilitou a compreensão e leitura por parte do público.

¹⁹⁸ Disponível em: <http://blog.estadaodados.com/advinhe-qual-partido-tem-mais-deputados-com-problema-na-justica/>. Acesso em 12 de fevereiro de 2016.

¹⁹⁹ <http://congressoemfoco.uol.com.br/noticias/os-deputados-sob-investigacao-no-supremo/>

²⁰⁰ O Infogr.am é como se fosse um Adobe Illustrator para a produção rápida e fácil de infográficos. Ele permite que qualquer pessoa (dedicada) e com um mínimo de experiência (ou curiosidade para aprender), consiga criar infografias em torno de três passos: selecionando um design, adicionando os dados e compartilhando ou incorporando a infografia. Outras informações em: <https://infogr.am/pt-br>

Figura V. 18 – Gráfico interativo traz deputados com problemas na justiça.

Fonte: blog.estadaodados.com, 2016

De acordo com Toledo, nem tudo rende uma ferramenta em visualização. No caso do Basômetro, por exemplo, faz sentido você ter uma ferramenta de acompanhamento permanente pois “é um filme de algo que continua acontecendo, tem um passado, um presente, um futuro. É uma situação mutante, ao longo do tempo, muda de um dia para o outro” (TOLEDO, 2016, anotações). Já no caso da Farra do Fies, o propósito era outro, o de narrar uma história. “o que queríamos era denunciar, contar uma historia. Então não foi necessário desenvolver uma ferramenta, que é algo mais complicado, cruzar várias bases de dados, que precisa de uma interpretação diferente” (ibidem, 2016).

Testes, adaptações e autodidatismo

No cenário Big Data, o bloquinho, a caneta, o gravador e a câmera, por exemplo, dividem espaço com recursos digitais como *Excel*, *Python*, *MySQL*, etc. Na Internet

há muitos dados de instituições onde a escala de investigação é maior; dados abertos, como do Instituto Brasileiro de Geografia e Estatística (IBGE), da Organização das Nações Unidas (ONU), do Banco Mundial, entre muitos outros. No Brasil, principalmente a partir da Lei de Acesso à Informação (LAI), de 2011.

O jornalismo de dados é um jeito diferente de apurar. Você não vai pra rua, faz fala-povo, usa métodos mais científicos. O prazer está em ler tabela, cheias de números, de dados até tirar conclusões (BURGARELLI, 2015, anotações).

Como todo o trabalho desenvolvido no Blog é experimental, são feitos muitos testes. Desde o uso de determinados infográficos em detrimento de outros e até de navegabilidade. Se as peças conseguem responder as principais perguntas que podem ir surgindo no decorrer da leitura; se os pontos trabalhados estão claramente compreensíveis. Esta avaliação fica por conta da própria equipe do Blog com leituras coletivas do material antes de serem publicados. De acordo com Toledo (2016), nunca foram feitas mudanças ou melhorias em reportagens publicadas no Blog, apenas desdobramentos do conteúdo postado. O que difere, por exemplo, das ferramentas e dos projetos permanentes. Como são constantes, às vezes são agregadas novas funcionalidades, melhorias na navegação. Além da alimentação com novas bases diante das recentes votações, como no caso do Basômetro.

A ferramenta Vai Passar²⁰¹ também é um exemplo. A proposta visa, a partir de um líder partidário, saber qual a chance de um projeto passar na Câmara. A pesquisa é feita com base na votação tendo as opções de maioria simples, maioria absoluta ou emenda à Constituição. “Pedimos ajuda do sistema de arte para fazer o designer, mas toda a programação foi feita por nós próprios”, comentou Toledo. O que nos leva a uma outra questão: ao autodidatismo da equipe.

Os profissionais hoje envolvidos diretamente no Blog tem a característica nata do autodidatismo. Tem como hábito a leitura de tutoriais e a busca por diferentes formas de contar histórias. Em vários trabalhos busca-se o que já foi feito em outros veículos de grande mídia, principalmente de fora do país, onde os núcleos de jornalismo de dados são mais desenvolvidos.

²⁰¹ Disponível em: <http://estadaodados.com/vai-passar/>

Figura V. 19 – Ferramenta interativa aponta qual a chance de um projeto passar na Câmara

Fonte: blog.estadaodados.com, 2016

Por exemplo, na reportagem *Compare o desempenho das seleções da Copa em 10 indicadores*²⁰² foi feito acompanhamento histórico do campeonato brasileiro inspirado num gráfico que foi publicado nos EUA, pelo estatístico Nate Silver (voltaremos a falar dele mais a frente), autor do blog FiveThirtyEight e que trabalha na ESPN. A ideia veio de fora, mas as ferramentas e os códigos foram desenvolvidos pela equipe brasileira.

²⁰² <http://blog.estadaodados.com/compare-o-desempenho-das-selecoes-da-copa-em-10-indicadores/>

Figura V. 20 – Infográfico interativo permite comparar o desempenho das seleções da Copa

COPA do MUNDO em Números

Ranking das seleções

A prova de que só tem um indicador que importa no futebol

Fonte: blog.estadaodados.com, junho de 2016

Faro jornalístico e formação das equipes

Mesmo trabalhando com números e dominando melhor recursos digitais, o coordenador José Roberto de Toledo aponta que é preciso o faro jornalístico. Ele defende a necessidade de um conjunto de características que são básicas de um bom repórter. Não somente curiosidade e paciência para cruzar informações e avaliar o que é notícia, mas principalmente não se contentar com o primeiro retorno ao lidar com dados. Para ele, o faro jornalístico é uma das essências da profissão que não deve estar presente apenas em jornalistas, mas na equipe como um todo.

(...) a capacidade de ver uma notícia quando se deparar com ela. É algo difícil de se especificar...é saber que ali tem uma novidade, uma novidade de interesse público e que você poderia chamar a atenção das pessoas para aquele assunto. É um conjunto de fatores. Um raciocínio questionador, que não se deixe levar pelo que os psicólogos chamam de sistema 1, imediatismo intuitivo, do primeiro contato com as informações. Pra coisa

funcionar você vai precisar do sistema 2, que é mais lento, mas que consiga colocar em contradição o que vê, com as suas próprias referências, seu próprio trabalho (TOLEDO, 2016).

Partindo deste depoimento retomamos a questão da integração da equipe. **Da mesma forma que é relevante que jornalistas entendam o básico em estatística, bancos de dados, entre outros. É relevante que estatísticos, programadores e designers entendam fundamentos do jornalismo para que ocorram diálogos, integrações e melhorias nas peças.** Isso tem acontecido na prática diária, com a formação das equipes e as demandas cotidianas. Em âmbito educacional diversas iniciativas estão disponíveis, entre cursos on-line gratuitos, cursos de curta duração, extensão para que os interessados possam dar o ponto de partida nas discussões que envolvem jornalismo de dados, jornalismo investigativo, Big Data, visualizações.

Em uma breve pesquisa, feita em dezembro de 2014, haviam dezenas de cursos. Mooc's gratuitos como o *Doing Journalism With Data*, do centro de notícias *Data Drive Journalism*²⁰³ e o *Introdução à Visualização de Dados e Infografia*²⁰⁴, que foi organizado pelo Centro Knight para o Jornalismo nas Américas e pela Associação Nacional de Jornais (ANJ); a Escola de Dados²⁰⁵ ministrou o curso *Introdução ao Jornalismo de Dados*, na Universidade Federal do Rio de Janeiro e no ano seguinte na Universidade de São Paulo. O Instituto Internacional de Ciências Sociais (IICS)²⁰⁶ também vem realizando um curso de uma semana de duração como parte do Programa de Jornalismo de Dados e Visualização. Há cursos de 3 meses como o de *Jornalismo Guiado por Dados*²⁰⁷, da Escola Superior de Propaganda e Marketing (ESPM)²⁰⁸ e o *Comunicação e Análise de Dados na Era do Big Data*, da Fundação Armando Álvares Penteado (FAAP).

Nas universidades há laboratórios e grupos de pesquisa que desenvolvem a temática como o Labic, da UFES, já citado em capítulos anteriores. Não há um levantamento sobre a inclusão de disciplinas específicas no ensino superior. A dissertação do pesquisador Rafael Alves (2012), da Universidade Federal de Santa Catarina (UFSC), apontou que há disciplinas que já contemplam à infografia, com tópicos que

²⁰³ <http://datajournalismcourse.net/index.php>

²⁰⁴ <https://knightcenter.utexas.edu/>

²⁰⁵ <http://escoladedados.org/>

²⁰⁶ <http://www.iics.edu.br/portfolio-item/programa-de-jornalismo-de-dados-e-visualizacao/>

²⁰⁷ http://www2.espm.br/sites/default/files/pagina/jornalismo_guiado_por_dados.pdf

²⁰⁸ http://www2.espm.br/sites/default/files/pagina/jornalismo_guiado_por_dados.pdf

trabalham o design, jornalismo visual, estatística, animação, *softwares* gráficos e desenho, entre disciplinas obrigatórias e optativas, com nomenclaturas como “Planejamento Gráfico”, “Computação Gráfica”, “Estatística”, “Design Gráfico”, “Práticas Fotográficas aplicadas ao Design”, “Programação Visual” e “Produção Editorial”. É um assunto complexo, multidisciplinar que desafia o mercado e também o ensino do jornalismo. Por isso a dificuldade em formar equipes de dados. São atividades que possuem suas peculiaridades, que necessitam de diferentes demandas do profissional e de investimento em armazenamento, computadores que tenham *softwares* de processamento mais ágeis, com grande capacidade de memória, entre outras peculiaridades que abordaremos ao longo deste capítulo.

Na Universidade de Columbia, em Nova York, por exemplo, há programas projetados para fornecer aos alunos habilidades em Ciência da Computação e Jornalismo de Dados de forma a prepará-los para as novas carreiras de mídia digital em jornalismo. Os alunos saem com certificação dupla. Entre as disciplinas oferecidas estão *machine learning*²⁰⁹, análise algorítmica, computação e humanidades, entre outras²¹⁰. De acordo com Tiago Mali, que trabalha hoje na revista Época, sem os conceitos básicos como regra de três, porcentagem, cálculo e variações é impossível conseguir matérias interessantes a partir de dados. O profissional lamenta o atraso nas faculdades de jornalismo no Brasil, que raramente incorporam conceitos do jornalismo de dados na grade curricular ou oferecem aulas sobre apuração com o uso de Big Data.

“Quando eu dava o curso de jornalismo de dados na Associação Brasileira de Jornalismo Investigativo (Abraji) era comum gastar muito tempo voltando a conceitos mais básicos. Essas coisas, se você não souber, pode tirar conclusões falsas. Não estou falando de ciência espacial, não tem nada complexo. Vale resgatar aquilo aprendido no colégio” (MALI apud PORTAL IMPRENSA, 2015)²¹¹.

Afinados com as crescentes disponibilidades de dados, a alfabetização das pessoas envolvidas no jornalismo de dados possibilita, em parte (e veremos o porquê mais a frente), a disseminação e um uso mais apropriado dos conteúdos. Entre os conhecimentos mais básicos estão a probabilidade, conceitos de estatística,

²⁰⁹ A *Machine Learning* ou Aprendizado de Máquina é um método de análise de dados que automatiza o desenvolvimento de modelos analíticos.

²¹⁰ <http://www.cs.columbia.edu/education/ms/journalism>.

²¹¹ Trecho retirado da matéria Jornalistas usam novas ferramentas para filtrar boas pautas no universo do Big Data. Disponível em: http://www.portalimprensa.com.br/especialtecnologia/conteudo_tecnologia_09.asp. Acesso em 1 de março de 2016.

associações, matemática, questões referentes a chamada ciência dos dados²¹², muitas vezes noções sobre linguagem de programação, que não precisa ser de baixo nível²¹³, mas pelo menos a escrita em notação, navegação criptografada. Em resumo, um olhar mais direcionado para as necessidades específicas de se trabalhar com o jornalismo de dados e a visualização de dados.

Desafios do novo

Apesar das dificuldades econômicas que as empresas jornalistas enfrentam, para nós, inicialmente, o maior empecilho para se trabalhar com dados estava na formação dos profissionais. Na tríade tempo, dinheiro e pessoas, a falta de capacitação seria um dos grandes problemas. No entanto, ao longo desta pesquisa, percebemos que os outros dois fatores passaram a ser os mais desafiadores. Hoje não seriam nem tanto as expertises dos profissionais ou os recursos tecnológicos, mas a falta de tempo e dinheiro das empresas, uma vez que as apurações e produções podem demorar de horas a meses de dedicação. Os recursos tecnológicos existem e mão-de-obra, mesmo que ainda precise de aperfeiçoamento, está presente no mercado.

O desafio é o mesmo do jornalismo. É um negócio que consome muito tempo e consome dinheiro. Pagar equipe, informática, Internet. E que não necessariamente você vai ganhar mais dinheiro com isso. Você não vai ganhar mais em anúncio, talvez até renda menos. O problema é esse. O maior desafio é que você não tem um modelo econômico sustentável para isso. A não ser que o jornal decida investir nesse tipo de trabalho, pelo prestígio, por que está cumprindo uma razão social, mas seja pelo que for, certamente, não é pela grana. Por isso poucas equipes se dedicam ao jornalismo de dados. Ele é um pedacinho dentro desse problema (TOLEDO, 2015).

Ao falar dos mesmos desafios, dos desafios do jornalismo, não nos vem em mente somente a questão econômica desta crise, mas fundamentalmente a de valores, função social e razão de ser. Há “feeds” de tudo, pílulas informativas ganham cada vez mais espaço em redações que parecem correr atrás de um ineditismo, de um furo de reportagem cada vez mais difícil de ser feito em ambientes onde as pessoas já se

²¹² Como já apontado no primeiro capítulo, uma ciência que procura descobrir conhecimento a partir de uma quantidade grande e pesada de dados. Os resultados são muitas vezes utilizados para a tomada de decisões e para fazer previsões.

²¹³ Quanto mais próxima da linguagem de máquina, mais baixo é o nível da linguagem. Quanto mais intuitiva para programar, mais alto é o nível da linguagem. Existem três níveis: baixo (Ex: Assembly), médio (Ex: C, C++) e alto (Ex: Java, Pascal). Fonte: softblue.com.br.

apoderaram das trocas informativas em alta velocidade. Não é que se deva parar de fazer, mas o jornalismo não se resume somente ao factual diário. Pelo contrário, num cenário de excessos informativos a contextualização se torna extremamente valiosa para o público, mas como valorar isso financeiramente se falamos de instituições economicamente ativas? Algumas iniciativas vem adotando consórcios para que consigam ser desenvolvidas. Citamos uma delas no início do capítulo II (Swiss Leaks), mas este é um caminho, não há solução para todos. A problemática é imensa e de rumos incertos.

Retomando a proposta desta pesquisa, percebemos que a partir do operacional das redações de dados é possível destacar pontos a serem desdobrados sobre o jornalismo e a visualização de dados. Muitos já começaram a ser discutidos na área acadêmica e de mercado, mesmo que de forma mais tímida. São questões relevantes que se configuram como tendências e nos ajudam a compreender melhor com o que estamos lidando e as dimensões que esta área pode tomar. Estamos falando de: matérias que envolvem previsões; a repercussão dos conteúdos; o armazenamento e a recuperação dos dados (datalização); as soluções e validações das peças jornalísticas; a lógica da produção dos conteúdos.

Previsão

Alguns conteúdos expõem fatos que já aconteceram, que estavam ocultos, outros apresentam tendências. Para Toledo fazer previsões é um grande desafio para o jornalismo. A matéria *Descubra o que seu time ainda pode fazer no Brasileirão*²¹⁴, publicada em outubro de 2015 foi uma das primeiras a trabalhar com este viés no blog Estadão Dados. É uma das poucas também que não aborda a temática política. A matéria focou na probabilidade do desempenho dos times, a chance que cada agremiação tinha de ser campeã no campeonato brasileiro de 2015, assim como a chance de classificação na Libertadores da América e de rebaixamento. Com infográficos interativos o leitor podia navegar pelas diversas possibilidades a partir dos dados da 28ª rodada de jogos.

²¹⁴ <http://blog.estadaodados.com/descubra-o-que-o-seu-time-ainda-pode-fazer-no-brasileirao/>

Primeiro a gente calculou a probabilidade. É um modelo, que já vem sendo usado por outros veículos, quantas partidas em casa, fora de casa e diante disso qual a probabilidade de vitórias. E nós acertamos tudo. Não me lembro de termos errado nada. Usávamos o passado como parâmetro para o futuro (TOLEDO, 2016).

Exemplos conhecidos neste “ramo” vem das previsões do estatístico e escritor americano, atualmente editor-chefe do blog de FiveThirtyEight²¹⁵, da ESPN e correspondente especial da *ABC News*, Nate Silver. Ele desenvolveu um sistema para prever o desempenho de jogadores da *Major League Baseball*. Assim como previu com sucesso o resultado da eleição presidencial de 2012, nos Estados Unidos, em todos os estados com base num modelo matemático. Enquanto comentaristas, em sua maioria conservadores, diziam que a vitória seria do republicano Mitt Romney, Nate Silver defendeu que o novo presidente americano seria Barack Obama. A mesma metodologia utilizada para fazer as análises de intenção de voto, ele alega que podem ser usadas em coberturas de economia, esportes, ciência e cotidiano (as mesmas editorias que estão presentes no blog FiveThirtyEight). Os métodos por ele adotados não se destinam a substituir o jornalismo convencional e a apuração de notícias. O objetivo de usar a estatística no jornalismo é a exploração de diferentes maneiras onde o consumidor pode usar dados para esclarecimentos.

Nate Silver lançou um Manifesto do Jornalismo de Dados²¹⁶, em 2014, onde defende a ideia de um jornalismo que tenha a estatística como elemento fundamental na reportagem. Para ele a qualidade da informação jornalística é importante, uma vez que boa parte da população se orienta pelo que é divulgado por jornalistas. Silver também defende que é preciso agregar valor ao jornalismo de dados e há potencial e mercado para isso. As pessoas deixam rastros de dados e muitos estão disponíveis de forma gratuita ou a preços baixos.

(...) eu nunca teria lançado FiveThirtyEight em 2008, e também não ampliaria sua cobertura agora, se não houvesse alguma necessidade para ele no mercado. Na minha opinião, as organizações de notícias convencionais, no seu conjunto, estão carentes de habilidades de jornalismo de dados. Parte disso é uma questão de auto-seleção. Os estudantes que entram na faculdade com a intenção de especialização em jornalismo ou comunicação têm pontuações acima da média em leitura e escrita, mas pontuações abaixo da média em matemática. Além disso, os jovens com habilidades

²¹⁵ <http://fivethirtyeight.com/>

²¹⁶ <http://fivethirtyeight.com/features/what-the-fox-knows/>. Acesso em 23 de fevereiro de 2016.

matemáticas fortes normalmente terão mais alternativas para o jornalismo quando eles embarcarem em suas carreiras e possam entrar em outros campos (SILVER, 2014, tradução nossa)²¹⁷.

Os dados tem uma origem, precisam ser apurados, interpretados e isto pode ser feito por um cientista ou por um jornalista, por exemplo. No entanto, é preciso chegar a conclusões e embasamentos consistentes. Sendo importante pensar em todo o processo para se fazer as perguntas certas aos dados. Para isso Nate Silver usa um processo dividido em quatro etapas: coleta; organização; explicação; generalizações e conclusões (SILVER, 2014)²¹⁸.

1^a) O primeiro passo é a coleta, a reunião de dados ou provas. Para um jornalista tradicional, isso envolve uma combinação de entrevistas, pesquisa documental e observações por parte do repórter. No jornalismo de dados também há outros caminhos de coletar informações, como encomendando pesquisas e realizando raspagem de dados em sites²¹⁹;

2^o) A segunda etapa é a organização. O jornalismo tradicional já tem estabelecido um meio de organização da informação: a formulação de notícias. A história é desenvolvida por ordem cronológica, por ordem de importância (a chamada pirâmide invertida) ou de inúmeras outras maneiras. Os jornalistas de dados, entretanto, organizam as informações utilizando estatísticas, colocando os dados em um banco de dados relacional ou construindo uma visualização de dados. Tendo ou não uma imagem, estas abordagens contam histórias de forma aprofundada²²⁰;

3^o) A terceira etapa é a explicação. Em termos jornalísticos, é preciso ir além das perguntas tradicionais: o que, onde, quando, como e porquê. É preciso ser mais do

²¹⁷ “Still, I would never have launched FiveThirtyEight in 2008, and I would not have chosen to broaden its coverage so extensively now, unless I thought there were some need for it in the marketplace. Conventional news organizations on the whole are lacking in data journalism skills, in my view. Some of this is a matter of self-selection. Students who enter college with the intent to major in journalism or communications have above-average test scores in reading and writing, but below-average scores in mathematics. Furthermore, young people with strong math skills will normally have more alternatives to journalism when they embark upon their careers and may enter other fields” (SILVER, 2014).

²¹⁸ Presente no Manifesto do Jornalismo de Dados disponível em: <http://fivethirtyeight.com/features/what-the-fox-knows/>. Acesso em 23 de fevereiro de 2016.

²¹⁹ “The first step is the collection of data or evidence. For a traditional journalist, this is likely to involve some combination of interviewing, documentary research and first-person observation. But data journalists also have ways of collecting information, such as by commissioning polls, performing experiments or scraping data from websites” (SILVER, 2014).

²²⁰ “The next step is organization. Traditional journalists have a well-established means of organizing information: They formulate a news story. The story might proceed chronologically, in order of importance (the inverted pyramid) or in some other fashion. Data journalists, meanwhile, can organize information by running descriptive statistics on it, by placing it into a relational database or by building a data visualization from it. Whether or not a picture is worth a thousand words, there is value in these approaches both as additional modes of storytelling and as foundations for further analysis” (SILVER, 2014).

que um jornalismo explicativo, descritivo. O jornalismo de dados possui cada vez mais um conjunto próprio de técnicas, executando, principalmente, vários tipos de testes estatísticos para procurar relacionar os dados²²¹;

4^a) A quarta etapa é das generalizações e conclusões. Para Nate Silver é difícil fazer previsões, mas uma das vantagens potenciais do jornalismo de dados é que ele generaliza melhor do que as abordagens tradicionais. Possui um embasamento em um conjunto de dados. O desafio é como fazer este jornalismo vivo, acessível a um grande público, sem sacrificar o rigor e a precisão. A dificuldade também está na não-linearidade do processo. É preciso ir e vir muitas vezes aos dados, mudar a organização deles.

Repercussão do conteúdo

Para Toledo, entre os desafios de se trabalhar com jornalismo de dados, visualizações, está a questão da imprevisibilidade da aceitação. Não há uma fórmula de sucesso. Os conteúdos e designs são trabalhados e levados aos leitores que podem ou não se apropriar das informações. Não há como saber previamente, mas há como acompanhar posteriormente, rastrear parte dos desdobramentos. Nos primeiros 20 meses do Basômetro, por exemplo, foram 17 reportagens produzidas somente no jornal O Estado de S. Paulo que abordaram os dados obtidos pela ferramenta. De certa forma, as peças jornalísticas funcionaram como fonte estruturada, que por si só são narrativas e geram narrativas.

Para a equipe do Blog, os retornos, as participações dos leitores são consequências, onde são sempre buscadas melhorias. No entanto, parece ainda ser colocado em segundo plano a questão das métricas de acompanhamento aos desdobramentos. As mensurações não estão presentes. Talvez porque ainda não haja uma real necessidade de acompanhamento, de justificativa de execução do projeto como um todo. Outras formas de reconhecimento relevantes, no entanto, tem acontecido como a própria premiação da matéria Farra do Fies, na 60^a edição do ExxonMobil 2015.

²²¹ “The third step is explanation. In journalistic terms, this might mean going beyond the who, what, where and when questions to those of why and how. In traditional journalism, stories of this nature are sometimes referred to as “news analysis” or “explanatory journalism.” Data journalists, again, have their own set of techniques — principally running various types of statistical tests to look for relationships in the data” (SILVER, 2014).

Concordamos com ANDERSON, BELL & SHIRKY (2012) e com a lógica do sistema narrativo proposto por Daniela Bertocchi sobre a importância deste acompanhamento.

Há uma relação estreita e simbiótica entre redes de usuários e jornalistas e dados. Os jornalistas devem ser capazes de analisar os dados e métricas que acompanham o seu próprio trabalho, têm uma familiaridade com a ideia de que as métricas representam atividade humana. Eles também devem saber como entender esse feedback e interpretá-lo de forma sensata, para que eles possam melhorar o alcance e conteúdo de suas histórias (ANDERSON, BELL & SHIRKY, 2013, p. 49).

Um ponto curioso levantando por Toledo é a veracidade que os conteúdos oriundos de grandes bases de dados apresentam. Mesmo por especialistas, parecem ganhar o status de verdade inquestionável; “as pessoas não duvidam tanto de matérias fundamentadas em tantos números”, disse. O Blog mesmo nunca foi acionado juridicamente por nenhum assunto abordado.

Armazenamento e recuperação de dados (dataização)

Os desafios no jornalismo são inúmeros e envolvem diferentes frentes de trabalho, que pela demanda, envolvem áreas correlatas e esforços que devem estar cada vez mais integrados nos próximos anos. O Blog seria um local de buscas constantes, fontes para outros jornalistas e pesquisadores? Ao serem perguntados sobre isso, Toledo e Burgarelli acreditam que não sejam realmente uma fonte permanente de consultas para jornalistas. Apesar de algumas repercussões, não acreditam que sejam vistos como um tipo de acervo pelos companheiros de profissão. **Outro ponto notado é que os conteúdos postados no Blog começam a ser difíceis de serem localizados. No caso da ferramenta *Vai Passar*, citada em entrevista por Toledo, só conseguimos localizá-la pela busca e não na seção em que se encontra, que é a Gráfico do Dia, sob o título “É possível prever as votações da Câmara – e isso diz muito sobre os nossos partidos”²²². Este fator nos despertou interesse sobre a questão do armazenamento e resgate dos conteúdos. Em um momento em que há poucos arquivos no Blog, fica mais fácil lidar com a recuperação das matérias, mas e depois? O Estadão Dados possui um servidor próprio, uma escolha em um primeiro momento de ordem operacional, mas que de acordo com Toledo não**

²²² <http://blog.estadaodados.com/vai-passar/>

interfere nas produções dos conteúdos. “Essa é uma questão de facilidade e custo, mas não é uma questão fundamental. Pode mudar a qualquer momento. São adaptações, melhorias. Não é um fator determinante para a realização ou não dos trabalhos desenvolvidos” (TOLEDO, 2016). E isto também está mais atrelado a suportabilidade de armazenar os grandes bancos de dados, que não necessariamente ficam disponíveis para os leitores.

Conversamos com Edmundo Leite, coordenador do Acervo²²³, do Jornal O Estado de S. Paulo, não especificamente sobre a questão do blog Estadão Dados, mas com o objetivo de que ele nos trouxesse informações sobre a recuperação de conteúdos no ambiente digital como um todo, o que achamos que será um caminho muito precioso diante do grande volume de dados na *web* e uma forma de cercar o objeto de estudo desta pesquisa. Em um primeiro momento nos deparamos com a informação de que o Brasil ainda não conta com uma instituição ou legislação que determine que conteúdos produzidos na *web* sejam vistos como patrimônio cultural e, desta forma, que sejam coletados, catalogados e armazenados. Por exemplo, as manifestações de junho de 2013²²⁴ estão arquivadas e garantidas segundo estes parâmetros apenas nos formatos impressos. O que predomina nos conteúdos *web* é um caráter mais efêmero, com sites que são constantemente atualizados ou que deixam de existir. Este apagão da memória da Internet já apresenta nuances. Contar algumas histórias são difíceis com a saída e atualizações recorrentes das páginas. Por exemplo, dificilmente seria possível contar a história e analisar as eleições presidenciais a partir dos sites dos candidatos. De acordo com matéria do Acervo, “desde 1998, já se foram quatro eleições presidenciais com a presença da Internet e nada foi coletado e sistematizado”²²⁵ (ENTINI, 2013).

O problema não é só do Brasil. Poucos países têm política ou instituições voltadas para o arquivamento web. Algumas iniciativas tem sido tomadas para minimizar o apagão. A mais antiga delas é o Internet Archive, de 1996. Através do Waybackmachine a instituição tem armazenado 347 bilhões de URLs de cerca de 40 países, inclusive o Brasil com 2,5 bilhões de capturas. No fim de 2012, o conteúdo total representava 10 petabytes, informa Kristine Hanna, diretora do Internet Archive, em

²²³ O Acervo responde hoje pelo armazenamento de todo conteúdo digital e não-digital do jornal. Disponível em: <http://acervo.estadao.com.br/>

²²⁴ https://pt.wikipedia.org/wiki/Protestos_no_Brasil_em_2013

²²⁵ ENTINI, Carlos Eduardo (2013). Quem arquivará a Internet para as futuras gerações? Disponível em: <http://acervo.estadao.com.br/noticias/acervo,quem-arquivara-a-internet-para-as-futuras-geracoes,9215,0.htm>. Acesso em 22 de fevereiro de 2016.

entrevista ao Estadão Acervo. Em 1996, foi a vez da Austrália coletar o conteúdo produzido e que fazia referência ao país. O exemplo foi seguido pela Suécia, no ano seguinte. Hoje, as instituições pioneiras estão reunidas no Consórcio Internacional de Preservação da Internet²²⁶ (ENTINI, 2013).

Todo o trabalho desenvolvido pela equipe do Acervo, incluindo ferramentas e treinamentos para os jornalistas é muito rico e interessante, mas um ponto chamou atenção na conversa e acompanhamento das atividades desenvolvidas, foi o termo “datalização”. Expressão que sintetiza o desafio dos profissionais e instituições que trabalham com preservação. Em matéria publicada em 2013, na página do Acervo, Edmundo Leite registrou o assunto na cobertura do III Seminário Internacional Arquivos e Museus de Pesquisa²²⁷, realizado em São Paulo. Nele a professora Marcia Lei Zeng, da Faculdade de Biblioteconomia e Ciência da Informação da Universidade de Kent, nos Estados Unidos apontou que a datalização é “a chave para que a enorme massa de conteúdos digitais possa ser organizada de modo a ser encontrada, compartilhada e integrada com outras bases de dados” (LEITE, 2013)²²⁸. Ela apontou ferramentas, práticas e protocolos disponíveis na *web* como o *Linked Open Data*²²⁹ e as buscas semânticas como meios para que documentos digitais possam ser encontrados com mais facilidade e se integrar a outros.

Uma página na Internet que estiver datalizada com a codificação e classificação corretas de metadados (informações sobre o conteúdo) poderá capturar automaticamente informações de outros sites que acrescentem mais dados àquele conteúdo originalmente isolado. Assim, uma instituição que possui uma carta digitalizada do imperador brasileiro Dom Pedro II poderá agregar ao documento dados de livros publicados sobre ele ou por ele, o verbete da Wikipédia ou qualquer outro site que possuir documentos de sua autoria, bem como incorporar um mapa on-line que tiver listado lugares onde ele nasceu ou pelo qual passou. Na mão inversa, um pesquisador que pesquisar sobre Dom Pedro II em outro site poderá chegar à carta digitalizada daquela instituição através dessas outras fontes, e não somente pelo próprio site da instituição ou dos buscadores convencionais disponíveis na internet (LEITE, 2013)²³⁰.

²²⁶ O Consórcio Internacional de Preservação da Internet (IIPC)²²⁶ é uma organização virtual, colaborativa, descentralizada, fundada em 2003. Com o objetivo de coletar, preservar e tornar acessível o conteúdo *web* para gerações futuras, atua na construção de tecnologias e no conhecimento em meios de armazenar sistematicamente as informações produzidas na *web*. Reúne cerca de 40 instituições (bibliotecas, arquivos, Internet Archive) em cerca de 30 países.

²²⁷ http://www.sescsp.org.br/programacao/seminario/3017_III+SEMINARIO+INTERNACIONAL+ARQUIVOS+DE+MUSEUS+E+PESQUISA#/content=apresentacao_mais

²²⁸ LEITE, Edmundo. Datalização é o desafio na era da digitalização. Disponível em: <http://acervo.estadao.com.br/noticias/acervo,datalizacao-e-o-desafio-na-era-da-digitalizacao,9285,0.htm>. Acesso em 5 de dezembro de 2015.

²²⁹ <https://vimeo.com/36752317>

²³⁰ Disponível em:

http://www.sescsp.org.br/programacao/seminario/3017_III+SEMINARIO+INTERNACIONAL+ARQUIVOS+DE+MUSEUS+E+PESQUISA-/content=apresentacao_mais. Acesso em 29 dezembro de 2015.

A datalização aparenta ser um procedimento que impactará em diversas áreas, inclusive no jornalismo. Além do jornalista escrever pensando que o texto será um dado para diversos formatos, o que reforça o já citado no capítulo II, a produção e armazenamento de informação deve ser compatível para qualquer fim. A recuperação diante do volume de dados é um ponto que estará a frente de discussões e melhor aplicabilidade, assim como a escrita em notação, por exemplo. Como meio de permanência, Ramalho (2015) recomenda uma escrita em notação, uma vez que é difícil definir quanto tempo um processador de texto como o *Microsoft Word* existirá, por exemplo. Será que daqui há 100, 200 anos ele estará em uso, em larga escala? Dispositivos poderão existir para fazer as traduções para outros *softwares*, mas talvez seja algo que deva ser melhor discutido por quem trabalha e produz informação: arquivamento e recuperação de dados.

Solução e validação das peças

A análise automática de grandes corpos textuais criou a possibilidade para os estudiosos de analisar milhões de documentos em vários idiomas com intervenção manual muito limitada. Entre as tecnologias e técnicas facilitadoras de análise encontram-se *machine translation*²³¹, categorização de tópicos e *machine learning*. Elas facilitam processos. No caso da *machine learning*, o uso de algoritmos permite o aprendizado de máquina de forma interativa, onde podem ser encontrados *insights* ocultos sem serem explicitamente programados para procurar algo específico.

O aspecto interativo do aprendizado de máquinas é importante porque, conforme os modelos são expostos a novos dados, eles são capazes de se adaptar de forma independente. Eles aprendem com os cálculos anteriores para produzir decisões e resultados confiáveis e reproduzíveis. É uma ciência que não é nova, mas que está ganhando um novo impulso (SAS, 2015)²³².

O diferencial da aprendizagem de máquina na contemporaneidade está na aplicação automática de cálculos matemáticos complexos a Big Data de forma cada vez mais rápida. Para Toledo, *machine learning* basicamente é fazer o algoritmo para que ele se

²³¹ A *Machine Translation* ou Tradução Automática (TA) é o processo automático de tradução de um idioma para outro através do computador. Em um nível mais básico, executa a substituição simples de palavras, mas em sistemas mais complexos podem reconhecer variações linguísticas e semânticas.

²³² Disponível em: http://www.sas.com/pt_br/insights/analytics/machine-learning.html. Acesso em 28 de dezembro de 2015.

auto-aperfeiçoe. Há inúmeras ferramentas gratuitas que oferecem serviços de soluções semelhantes. A IBM, por exemplo, desenvolveu o *Watson*, um sistema desenhado para responder às perguntas de um dos programas de televisão americanos mais importantes chamado *Jeopardy!*. O *Watson* é pioneiro no processamento da linguagem e analítica e, a partir dele, foram divulgadas 18 ferramentas gratuitas em plataforma de nuvem²³³, como o *Visual Recognition*, que permite analisar imagens ou trechos de vídeos para compreender o que acontece em uma cena²³⁴; *Cognitive Commerce*, que traz oportunidades comerciais para consumidores e negócios por meio de recomendações mais relevantes, de forma personalizada²³⁵ e o *Cognitive Graph*, que é um recurso que agrupa conhecimento fornecido por terceiros, dados internos, fontes privadas de informações e usa modelos de domínio específico em respostas com representação gráfica²³⁶. No caso específico do Estadão Dados são usados em algumas peças os conceitos de machine learning.

Você trabalha algumas coisas em R pra desenvolver o cálculo estatístico. Por exemplo, a cada rodada do campeonato brasileiro a probabilidade de um time ganhar é maior do que a de outro e a máquina já faz o cálculo automaticamente. No caso da matéria sobre o campeonato brasileiro, quem fez a alimentação dos dados foi o Guilherme, assim como o modelo estatístico. Ele pois a mão na massa e o meu papel foi o de ser o chato, criticar, pedir para mudar algoritmo (TOLEDO, 2016, anotações).

Muitas vezes as máquinas fazem os cálculos, mas a checagem das informações acontece como em qualquer matéria. Como comenta o coordenador do blog Estadão Dados, com eles não há um protocolo fixo de procedimentos, uma vez que os casos são variáveis. “Buscamos ver se o resultado faz sentido, se as somas realmente dão 100%. Não há uma fórmula, um protocolo de checagem. A gente trabalha com assuntos tão variados que prevalece a preocupação de não gerar nenhum processo” (TOLEDO, 2016, anotações).

O importante é saber exatamente o que se quer e como trabalhar os dados. Compreender que há diferenças entre aprendizado de máquina e mineração de dados,

²³³ Disponível em: <http://idgnow.com.br/ti-corporativa/2016/02/26/watson-oferece-18-ferramentas-gratuitas-em-plataforma-de-nuvem-da-ibm/>

²³⁴ Usando recursos de aprendizado de máquinas, um sistema ajuda no reconhecimento e classificação semântica de informações visuais, com objetos ou eventos. A ferramenta vem com alguns modelos de imagens já reconhecidos para prover resultados com base em análise e aproximação. A versão beta é gratuita.

²³⁵ O serviço, disponibilizado no Bluemix por um parceiro IBM chamado *Cognitive Scale*, ajuda empresas em transações comerciais. Também é gratuito em versão beta.

²³⁶ A solução usa *machine learning* e pode ser aplicada em diferentes tipos de problemas. Grátis em versão beta.

por exemplo, mas que há muitas outras complexidades do processo que interferem no resultado final dos dados obtidos. Desde questões mais simples como a compatibilidade de máquinas, quanto ao uso de cálculos de forma correta. De qualquer forma, estas variações devem gerar somente pequenas variações. No início do livro *Análise Política & Jornalismo de Dados* (TOLEDO et. al, 2014), é feito um alerta sobre possíveis variáveis no processamento da base de dados da ferramenta Basômetro conforme o navegador do usuário.

Isso significa que podem haver diferenças mínimas nos resultados dos filtros e cruzamentos, em função das características do computador de cada um. Isso não invalida, com base em diversos testes, os resultados colhidos, e não ofusca o brilho da ferramenta, que nos permite compreender e analisar a realidade de nosso parlamento federal em anos recentes (TOLEDO et al. 2015, p. 12).

Novamente, o importante e relevante é que as variações sejam reconhecidas e mínimas, que permitam que o usuário faça idas e voltas e tenha conhecimentos sobre as fontes dos dados e caso interesse, cheque as informações, a partir das fontes originais.

A lógica da produção dos conteúdos

No blog Estadão Dados não há uma rotina de trabalho propriamente dita como nas demais redações jornalísticas, inclusive como a do próprio jornal O Estado de S. Paulo. De acordo com Rodrigo Burgarelli, o processo de produção da pauta começa com uma pergunta, seguida de uma parte mais técnica que é a obtenção da base de dados, “tratar, fazer a análise, algo que, muitas vezes, não vai ser feito por quem teve a ideia. E depois tem que entrevistar pessoas, fazer fotos, vídeos, escrever. É uma tarefa multidisciplinar” (BURGARELLI apud PORTAL IMPRENSA, 2015). São diversas frentes de trabalho desenvolvidas e a atualização do Blog não tem a obrigatoriedade de ser diária. O gerenciamento é complexo.

A gente está sempre caçando novas bases de dados e olhando as antigas com novos pontos de vista. As pautas, as ideias, vão surgindo. Você às vezes perde tempo com uma base e ela não te leva a lugar nenhum. A gente faz, sem uma ideia pré-concebida, de espírito aberto, para tentar buscar os assuntos. O método é um pouco caótico, eu diria. Não é sempre do mesmo jeito. Na verdade a gente tem muito mais ideia do que capacidade de executar, mas assuntos não faltam. O que falta é tempo (TOLEDO, 2016, anotações).

As lógicas dos processos de produção interessam a presente pesquisadora, que em outros momentos já foi buscar conhecimentos básicos sobre gerenciamento de projetos e desenvolvimento de *softwares*. Assunto que acabou cruzando com as informações obtidas em setembro de 2015, quando o programador e designer Carlos Lemos, que já fez parte da equipe do blog Estadão Dados e hoje integra a equipe do Globo Esporte fez uma palestra no encontro Hacks Hackers denominada “A cultura dos desenvolvedores dentro da redação do Globo Esporte”²³⁷. O objetivo da palestra foi discutir exatamente a função do programador e da programação dentro das redações jornalísticas. Se antes a única função era solucionar *bugs* no sistema, agora, “com o fim do *Flash*, as luxuosas visualizações de dados e a evolução natural da produção de conteúdo para a rede nos mostram que precisávamos de equipes mais robustas dentro das empresas jornalísticas”, disse Lemos (2015, anotações). A pergunta-problema da palestra foi: como se deu essa integração entre programadores e redação jornalística?

Carlos Lemos apontou qual o seu trabalho hoje dentro do ambiente redacional. Um cotidiano que envolve o manuseio de grande volume de dados; armazenamento e curadoria dos dados aplicado nas editorias jornalísticas; integração de diferentes profissionais na execução de cada peça jornalística; a necessidade de alimentação atualizada, mas não diária. Elementos que temos discutido e apontado ao longo desta pesquisa. No entanto, o ponto que chamou a atenção é um desdobramento da questão dos tempos de produção e execução diferentes em cada etapa do processo da construção narrativa e da necessidade das expertises dos profissionais envolvidos estarem integrados. Lemos defende e aplica o uso de técnicas de planejamento e gestão. No caso específico apresentado durante a palestra, do uso de um *Scrum*²³⁸ jornalístico para se trabalhar com dados.

O *Scrum* é um processo de trabalho em equipe, em um projeto que possui uma mesma finalidade, desenvolvimento de tarefas e metas que são compartilhadas e possuem um *feedback* (retorno) rápido de produtos e dos processos de trabalho. É um framework

²³⁷ O bate-papo com Carlos Lemos aconteceu no dia 24 de setembro de 2015, no Instituto Internacional de Ciências Sociais (IICS), em São Paulo. Disponível em: http://www.meetup.com/pt-BR/hackshackers/events/225232624/?eventId=225232624&chapter_analytics_code=UA-30964380-17.

²³⁸ O gerente de desenvolvimento de *softwares* Hamid Shojaee, fundador da empresa Axosoft fez um vídeo de 10 minutos explicando de forma ilustrativa o que é um scrum. Ele trabalha com *softwares* de gerenciamento de projetos ágil usado por mais de 11.000 equipes em todo o mundo. Disponível em: <https://www.youtube.com/watch?v=XU0lIRltyFM>. Acesso em 29 de fevereiro de 2016.

(caixa de ferramentas) de desenvolvimento iterativo e incremental utilizado no gerenciamento de projetos e desenvolvimento de *software* ágil. Funciona como um conjunto de valores, princípios e práticas que fornecem a base para que a organização adicione práticas particulares de engenharia e gestão e que sejam relevantes para a realidade de cada empresa, fazendo com que cada uma tenha um *Scrum* próprio²³⁹. O foco está no gerenciamento de projetos onde é difícil planejar à frente. Não é um sistema que descreve o que fazer em cada situação, mas é usado para trabalhos complexos nos quais é impossível prever tudo o que irá ocorrer. Temos a visão macro de uma narrativa como sistema, um processo em fluxo, como discutido no capítulo III.

Agora estamos falando da aproximação entre os sistemas de informação das técnicas de comunicação ágil em várias escalas de uso. Não somente para dentro das redações enquanto integrantes de uma empresa, mas para dentro das editorias e também de cada processo narrativo. É um assunto um tanto quanto complexo, mas já em uso em uma etapa em que as redações ou grupos de trabalhos jornalísticos já estão inseridos num cenário de uso de base de dados, ou seja, que já passaram pela etapa de que jornalismo de dados se faz com equipes que em um único trabalho compreendem técnicas de extração de dados, usos de ferramentas e *softwares*, análises e correlação de dados. Trabalham com a visualização de dados e visualização da informação e precisam de alguma forma sistematizar os processos de execução dos conteúdos no ambiente interno diante do fluxo informativo dos processos de narração da contemporaneidade. O que nos leva a questão de que, enquanto jornalistas, dependendo das funções exercidas e da área de atuação, devemos atentar para gerenciar dados em uma lógica de sistemas de produção e execução de projetos.

Sustentação teórica vinculada a *core category*

A nossa *core category* para definir que a peça jornalística se trata de uma visualização de dados, aos moldes do que discutimos ao longo desta pesquisa, está na relação entre: o Big Data, o jornalismo de dados e o sistema narrativo.

²³⁹ Outras informações podem ser obtidas em www.desenvolvimentoagil.com.br.

Onde a visualização de dados jornalísticos aparece como um formato em constante diálogo com o design. Uma construção que acontece de forma não linear e que tem processos próprios que dependem dos dados, das linguagens, das ferramentas, dos locais de armazenamento e das formas de resgate dos conteúdos. Compreender porque o texto analisado (seja ele numérico ou não) se apresenta de determinada maneira é compreender o seu sistema de organização, a sua estrutura. Identificamos conceituações, mas não processos regulares e repetíveis de produção, uma vez que, como pode ser observado pelos exemplo, não há uma única maneira de execução para se trabalhar os conteúdos originários de bases de dados e que geram visualizações de dados. O nosso diagrama pode ser compreendido da seguinte maneira:

Figura V. 21 – A visualização de dados no jornalismo contemporâneo

Fonte: Mayanna Estevanim, 2016

Através dos exemplos apresentados ao longo deste capítulo sustentamos que o fator tecnológico é um dos pré-requisitos para o desenvolvimento das peças jornalísticas, auxiliando no manuseio de dados e metadados. No ambiente digital, a narrativa deixou de ser compreendida como texto e imagens estaticamente diagramados. É dinâmica e em constante transformação no tempo e no espaço digital, onde o usuário experimentou o fluxo contínuo de eventos. Quando a narrativa chegou a uma interface para o usuário ela já foi formatada, mas o seu fluxo continuou uma vez que os usuários puderam acrescentar comentários, compartilhar informações, curtir e sugerir melhorias. Os exemplos envolveram investigações e inúmeras correlações de dados que foram apresentadas aos leitores de maneiras distintas como em ferramentas, infográficos e mapas.

As bases de dados não necessariamente apareceram ao usuário/leitor, uma vez que serviram apenas como fonte para a produção de conteúdos jornalísticos, além de apresentações infográficas, mapas, entre outros recursos visuais. Confirmando que a visualização é um meio de transformar dados em informação semântica, criando ou não ferramentas para que qualquer pessoa faça a leitura e interpretação de um processo informativo. Os jornalistas envolvidos foram modeladores da narrativa em camadas, tendo o auxílio de equipes humanas e não-humanas resultando em uma experiência narrativa centrada nos usuários. Quanto maior a familiaridade com as camadas subterrâneas do sistema narrativo, melhores foram as formas de atuação, intervenção e de contar histórias no ciberespaço.

No entanto, o que ficou claro é que as novas possibilidades para o jornalismo exigem diferentes formas de organização e somente a incorporação de técnicas não serão suficientes para a adaptação a este novo ecossistema. Há um caráter técnico de escolhas de apresentação das peças, telas, bandas largas, políticas de uso de dados, mas também há questões que envolvem a essência jornalística na contemporaneidade. É preciso dar valor e funcionalidade ao jornalismo. Não são só produtos, são fluxos, onde o jornalismo esteve no centro da execução, independentemente da formação do profissional envolvido.

O nível de profundidade da pesquisa de dados nas matérias foi fundamental para a criação das peças premiadas, que além do ineditismo no cenário nacional, puderam responder perguntas, permitiram cruzamento dados e quando não, deram sentido a eles priorizando elementos visuais que foram funcionais e contribuíram para a eficácia do entendimento dos assuntos abordados.

Considerações

Neste capítulo partimos, inicialmente, de uma visão macro do blog Estadão Dados onde buscamos compreender as dinâmicas do trabalho da visualização de dados dentro da redação, atentando para a formação dos profissionais envolvidos, para as generalizações sobre construções noticiosas, estrutura de trabalho e investimentos no núcleo de dados; assim como as principais dificuldades em lidar com dados. Em um segundo momento optamos por nos aprofundar em exemplos de peças jornalísticas desenvolvidas pela equipe, com foco na ferramenta em visualização de dados Basômetro e na série de reportagens-denúncia Farra do Fies. Acreditamos que a partir deste caminho, com a seleção da amostra e com a sustentação teórica pudemos responder a pergunta-problema desta pesquisa: o que é a visualização de dados no jornalismo? Como ela se insere na prática e na narrativa do jornalismo de dados proporcionando diferenciais relevantes? As observações mais sistematizadas e as reflexões desencadeadas a partir do ambiente da redação jornalística nos geraram empiricamente categorias, sendo possível percorrer o processo de geração das mesmas. Acreditamos que as categorias e os processos apresentados abrem novas perspectivas sobre os estudos, sendo útil, principalmente, para os envolvidos diretamente no jornalismo de dados, sugerindo pistas para mudanças e inovação.

De posse do material de campo, entrevistas e observações sistematizadas, as reflexões foram desencadeadas sob dois aspectos: os pontos jornalísticos que se destacavam no ambiente redacional e os desafios de se trabalhar com a visualização de dados no jornalismo. Os pontos jornalísticos destacados foram:

- finalidade do conteúdo – que define o design que chegará ao leitor e o desenrolar da construção narrativa;

- os testes, adaptações e autodidatismo – as experimentações feitas por parte da equipe para que a mensagem seja passada e as possíveis indagações dos leitores sejam esclarecidas;

- o faro jornalístico e a formação da equipe – características básicas que devem nortear todos os envolvidos na produção jornalística para que as perguntas certas sejam feitas aos dados.

Entre os desafios para se trabalhar com o jornalismo de dados, que impactam diretamente nas tendências e continuidade da prática, destacamos a tríade tempo, dinheiro e pessoas, onde a falta de mão-de-obra qualificada não aparece como o maior empecilho, mas sim a falta de recursos para manter apurações e produções que ocorrem de maneira distinta da lógica predominante do exercício do jornalismo, num cenário já desfavorecido por modelos de negócios que não mais se sustentam. A partir do acompanhamento da redação, as reflexões desencadeadas levam a situações, tendências e possíveis desdobramentos do fazer visualização e jornalismo de dados, que são:

- os conteúdos jornalísticos de previsão – que utilizam métodos para abordar situações futuras não de forma especulativa, mas baseadas em dados;

- a repercussão das peças – sua imprevisibilidade e muitas vezes necessidades de ajustes na busca por melhorias na apresentação ou entrega dos dados;

- o armazenamento e recuperação de dados – onde o termo datalização surge como síntese do desafio dos profissionais e instituições que trabalham com a preservação. Fato este que os jornalistas ainda precisam ter mais zelo;

- a solução e validação das peças – usos de algoritmos e robôs em tecnologias facilitadoras de análises de grandes conjuntos de dados. A tão falada aprendizagem de máquina que auxilia a encontrar variáveis nos dados e nas construções das bases de dados;

- a lógica da produção dos conteúdos – onde estão presentes diferentes maneiras de se

trabalhar no ambiente redacional voltado para a prática do jornalismo de dados. Falamos aqui do gerenciamento que envolve o desenvolvimento de peças simultaneamente.

A visualização de dados jornalísticos para nós é um formato que tem processos próprios que dependem dos dados, das linguagens e ferramentas, locais de armazenamento e formas de resgate. A *core category* para definir que a peça jornalística se trata de uma visualização de dados baseia-se, portanto, na relação entre Big Data, jornalismo de dados, sistema narrativo, a partir do formato visualização de dados. Tendo um diálogo constante e não linear com o design das peças.

VI. CONSIDERAÇÕES FINAIS

No dia 8 de março de 2016, o jornal O Estado de S. Paulo esteve em 4º lugar no *Trending Topics* do *Twitter* Brasil. Como parte de uma campanha pelo Dia Internacional da Mulher, o jornal abriu seu perfil e divulgou relatos de violência exatamente como eles são registrados no país, a cada 7 minutos²⁴⁰. O Estadão registrou 11,6 milhões de visualizações em seu perfil. Um aumento de mais de 152% em sua média diária. Segundo dados do *Twitter*, em 24 horas a *hashtag* #7minutos1Denuncia teve 28,2 mil menções. As 205 denúncias foram retuitadas mais de 23 mil vezes. De acordo com o professor e pesquisador Fábio Malini, do Labic/UFES, foi um exemplo de jornalismo e transmidialidade. “Só isso evidencia que o “espalhar em massa”, ação de sujeitos diferentes (incluindo aí grupos feministas), alcançou o objetivo do Estadão em propagar a indignação contra a alta estatística de violência contra a mulher”²⁴¹, disse Malini.

A análise dos dados e geração da imagem gerada pela *hashtag* #7minutos1Denuncia foi divulgada no dia 9 de março, um dia depois do fenômeno. A ferramenta utilizada foi o Gephi, mas este exemplo envolveu um método de análise de redes que busca identificar, processar e interpretar os pontos de vistas que são expressos no espaço e tempo das interações em redes sociais. Pontos de vista que muitas vezes partem de reflexões complexas que envolvem teorias antropológicas (como as formuladas por Eduardo Viveiros de Castro de onde são retirados os conceitos de perspectiva e relação); os estudos sobre cartografia, grupos, mediadores, intermediários (que advém muitas vezes de Bruno Latour sobre a teoria ator-rede); envolvem conceitos de clusterização, modularidade, centralidade, densidade (oriundos da teoria dos grafos). Informações que dizem respeito sobre comportamento coletivo, engajamento do público e retorno imediato para o próprio veículo de comunicação sobre a sua iniciativa. É mais um exemplo que está inserido no universo que objetivamos apresentar ao longo desta pesquisa, que envolve um grande volume, variedade e velocidade ao lidar com dados. Envolve correlação de informações, análise,

²⁴⁰ Disponível em: <http://brasil.estadao.com.br/noticias/geral,campanha-do-estado-no-dia-da-mulher-repercute-no-twitter,10000020436>. Acesso em 10 de março de 2016.

²⁴¹ Disponível em: <https://www.facebook.com/photo.php?fbid=10154019269101151&set=a.477225711150.296154.703916150&type=3&theater>. Acesso em 9 de março de 2016.

apresentação visual para a melhor compreensão do fenômeno e da repercussão pelo público. Não focamos nesta pesquisa especificamente na análise de redes sociais, mas sim em um jornalismo que se apropria de recursos tecnológicos, cruzamento de banco de dados, métodos de análise, construções de mapas, cartografias, infográficos dinâmicos em narrativas diferenciadas que fornecem ao usuário o entendimento de temas complexos de relevância social, de forma visual, diante das diferentes possibilidades de compreensão informativa da sociedade contemporânea.

Figura 22 – *Hashtag #7minutos1Denuncia – Twitter, 8 de março de 2016*

Fonte: – Perfil Fábio Malini, Facebook, 2016²⁴²

O intuito desta dissertação foi problematizar a visualização de dados no jornalismo brasileiro partindo exatamente do que é a visualização de dados, mais especificamente no jornalismo de dados, e diante dos apontamentos sobre seu conceito e prática questionar como esta visualização se insere na narrativa jornalística proporcionando diferenciais relevantes. Por relevantes partimos da premissa de pautas de interesse

²⁴² Disponível em:
<https://www.facebook.com/photo.php?fbid=10154019269101151&set=a.477225711150.296154.703916150&type=3&theater>.
 Acesso em 9 de março de 2016.

público, que envolvem maior criticidade, maior aprofundamento e contextualização dos conteúdos no *Big Data*. Por ser um tema complexo da contemporaneidade, foram muitas idas e vindas para cercar o objeto conceitualmente e necessários diferentes olhares e reflexões teórico-metodológicas para entender a atualidade.

O quarto capítulo foi essencial para demonstrar os caminhos metodológicos percorridos nesta pesquisa e os rumos para as observações de campo que foram apresentadas no capítulo V. Sintetizou as dificuldades dos estudos em Comunicação na contemporaneidade e o olhar que demos aos fenômenos neste contexto. Um olhar que tem a Comunicação como centro, que perpassa todas as atividades comunicativas e informativas de forma transversal. Optamos por resiliências diante das abordagens teórico-metodológicas para que as mesmas conseguissem abarcar e sustentar as reflexões referentes ao dinâmico campo de estudos. Elementos que justificaram nossas escolhas metodológicas que envolveram cruzamentos de distintos caminhos: pesquisa bibliográfica, observação não-participante, estudo de caso, fundamentos da Grounded Theory, cursos, disciplinas, grupo de pesquisa Com+, palestras, estágio de um ano na disciplina Jornalismo on-line. Refletimos dos dados de campo para as teorias emergente e também fizemos a trajetória contrária para fazer as proposições e interpretações. Sustentamos o paradigma Jornalismo Digital em Base de Dados, tendo as contribuições dos conceitos de formato (RAMOS, 2012 e MACHADO, 2003), de jornalismo pós-industrial (COSTA, 2014), sistema narrativo e antenarrativa (BERTOCCHI, 2013) como meios de amadurecimento da compreensão do objeto proposto.

Nosso desafio nesta pesquisa foi o foco não no resultado final, mas na compreensão sobre como os processos se estabelecem no âmbito deste manuseio de dados e metadados; de tornar visível o invisível. Para nós a visualização de dados se confirmou como um processo não-linear, trabalhado no jornalismo de dados que se encontra em área fronteiriça de estudo, possuindo características próprias que o distingue de outras práticas. Consideramos a visualização de dados jornalísticos como um formato que recorre à utilização de *softwares* de processamento de dados que estão envolvidos nas camadas de dados e metadados do sistema narrativo, traduzindo a complexidade dos dados e suas correlações através de modelos visuais. Uma ação que envolve o jornalismo visual, científico, investigativo, as técnicas do RAC e que

emerge a partir de congruências e manuseio de grandes volumes de dados. Podemos dizer que neste ecossistema comunicacional o jornalista passa a assumir um papel de humanização dos dados servindo como um investigador, alguém que traduz e conta fatos. É também um curador de notícias diante da variedade, velocidade e quantidade de informações disponíveis.

Nossa preocupação teve como foco as discussões em âmbito nacional, uma vez que as questões políticas e leis de transparência interferem na produção e disseminação da prática no país. Distinguimos em três principais áreas como as visualizações de dados vem aparecendo no cenário jornalístico brasileiro: nos veículos tradicionais de mídia, nas iniciativas independentes e nos laboratórios acadêmicos. De imediato percebemos que a formação das equipes, o tempo de produção dos conteúdos e as fontes dos dados são distintos de outras práticas jornalísticas, frutos de um sistema narrativo que na contemporaneidade adquire novas rotinas produtivas. Sustentamos que o jornalismo de dados e a visualização de dados estão inseridos em sistemas narrativos condizentes com processos inovadores no jornalismo contemporâneo, onde o jornalista não é somente um narrador de fatos, ele é o responsável por coletar, apurar, interpretar e contextualizar as informações. Recorre a gráficos interativos, solicita imagens e informação ao público/leitor, disponibiliza acesso direto a bancos de dados e distribui matérias por redes sociais.

O uso de bases de dados reconfigura, pelo viés do cruzamento de dados, um conteúdo mais original ao jornalismo, onde os dados podem ser somados, comparados, integrados entre si. Neste contexto, as bases de dados são essenciais, mas não estão no centro do sistema narrativo, uma vez que devem ser considerados outros elementos como os algoritmos, os sistemas publicadores e os metadados. Para nós, os dados, metadados e formatos possuem importância similar no processo de construção narrativa e ignorar uma destas camadas representa perdas no jornalismo de dados e consequentemente na visualização de dados.

As possibilidades de formatação jornalística envolvem o agenciamento de inúmeras camadas, onde estes formatos não se constituem como gêneros jornalísticos, mas como “design de gênero” porque, enquanto o gênero discursivo é da ordem da língua, o formato é da ordem das linguagens modelizadas pelos códigos culturais

tecnológicos. Não estando no ambiente semiótico da codificação tecnológica, o formato não pode ser concebido. Seja como um processo ou como um produto, eles possuem uma lógica interna de construção. O formato é mais do que um artefato, do que um resultado final. Ele é um processo, em fluxo, resultado de um sistema aberto e que está em constante mudança sejam eles visualizações de dados, com ou sem usos de algoritmos em seu desenvolvimento.

Defendemos ao longo da pesquisa que o jornalista não foi substituído e muito menos extinto, mas sim, neste cenário, exerce funções de verificar, interpretar e dar sentido à enxurrada de textos, áudios, fotos, vídeos em narrativas originais e criativas. O olhar aos dados tem um caráter imersivo sobre determinado assunto, é de cunho investigativo e muitas vezes inclui outros meios, como entrevistas, depoimentos, elementos que venham a contribuir para o conhecimento. O jornalista pode participar tanto como produtor do conteúdo como do desenvolvimento de *softwares*. É um curador de notícias que se familiarizado com as camadas subterrâneas do sistema narrativo encontra melhores maneiras de contar histórias. A camada mais próxima da superfície é o formato. Eles modelizam as linguagens digitais sendo interfaces para as bases de dados disponíveis no ciberespaço. Na camada da antenarração é onde ocorrem grandes influências no sistema narrativo, que podem alterar seu comportamento pensando numa computação social centrada no humano. Pensar no formato narrativo digital é pensar em como os usuários se apropriam de serviços, tecnologias e dispositivos num contexto digital. E neste contexto, o jornalismo se desenvolve no agenciamento coletivo de diferentes atores (jornalistas, designers, engenheiros, programadores, arquitetos de informação, especialistas em negócios, usuários, algoritmos, robôs, *softwares*, entre outros).

Acreditamos que a pesquisa contribui para os estudos da Comunicação contextualizando as teorias com o cenário atual proporcionado pelos avanços tecnológicos na coleta e no auxílio ao tratamento de grandes volumes de dados no jornalismo. No capítulo V mostramos a construção do conteúdo noticioso a partir da visualização em base de dados, tendo como objeto de estudo o jornalismo de dados no Brasil, com estudo de caso do blog Estadão Dados, do jornal O Estado de S. Paulo. A escolha se deu por ter no Blog tanto conteúdos gráficos, quanto projetos em visualização, pelas atualizações mais constantes e por estar em um cenário brasileiro

de investimentos nesta área, o que nos possibilitou um estudo com maior embasamento de campo diante dos números de experiências em práticas. Buscamos com o estudo de caso, um método de coletar dados do corpus, a possibilidade de aprofundar a pesquisa, de avaliar quantitativamente e qualitativamente os processos envolvidos na produção de conteúdos que envolvam o manuseio de grande volume de dados. O estudo de caso permitiu a profundidade necessária para obter parte da compreensão desejada. A pesquisa foi prosseguida pela Grounded Theory, que fez emergir processos subjacentes às afirmações dos participantes e ao fenômeno observado.

Observamos como os conteúdos jornalísticos foram desenvolvidos a partir de fontes, uso de *softwares* e metodologias de análise de dados. E a partir desta fase, indagamos sobre o papel do jornalista neste cenário onde dados podem ser somados, comparados e integrados entre si, complementando os grandes blocos de texto que o computador só consegue mostrar, mas não correlacionar, sem ser programado para isso. A criação de programas e algoritmos que substituem o trabalho humano de reportagem envolve uma série de decisões que devem ser passíveis de explicação e responsabilização para todos os afetados. Aplicativos de *softwares* baseados na *web* precisam da mediação humana, profissional, para emergir e interpretar as informações. Não basta torná-las acessíveis ou legíveis; é preciso torná-las compreensíveis e relevantes, ter valor jornalístico. Entre os desafios de se trabalhar com o jornalismo de dados percebemos que a falta de mão-de-obra qualificada não apareceu como o maior empecilho, mas sim a falta de recursos para manter apurações e produções que ocorrem de maneira distinta da lógica predominante do exercício do jornalismo, num cenário já desfavorecido por modelos de negócios que não mais se sustentam.

A nossa *core category* para definir que a peça jornalística se trata de uma visualização de dados, aos moldes do que discutimos ao longo desta pesquisa, foi a relação entre: Big Data, Jornalismo de Dados, Sistema Narrativo, Formato, Visualização de Dados e Design. Identificamos conceituações, mas não os processos de produção padronizáveis, uma vez que, como pode ser observado, não há uma única maneira de execução para se trabalhar os conteúdos originários de bases de dados e que geram visualizações de dados. O acompanhamento da redação nos levou ao confronto entre teorias emergentes e as observações. Os pontos jornalísticos destacados foram: 1.

finalidade do conteúdo (que define o design que chegará ao leitor e o desenrolar da construção narrativa); 2. os testes, adaptações e autodidatismo (as experimentações feitas por parte da equipe para que a mensagem seja passada e as possíveis indagações dos leitores sejam esclarecidas); 3. o faro jornalístico e a formação da equipe (características básicas que devem nortear todos os envolvidos na produção jornalística para que as perguntas certas sejam feitas aos dados).

Entre as reflexões desencadeadas, que nos levaram a situações, tendências e possíveis desdobramentos do fazer visualização e jornalismo de dados, estiveram: 1. os conteúdos jornalísticos de previsão (métodos para abordar situações futuras não de forma especulativa, mas baseadas em dados); 2. a repercussão das peças jornalísticas (pela sua imprevisibilidade de aceitação e muitas vezes necessidades de ajustes na busca por melhorias na apresentação e entrega dos dados); 3. o armazenamento e recuperação de dados (os desafios no processo de arquivamento e resgate dos conteúdos. Fato que definitivamente os jornalistas precisam ter mais zelo); 4. a solução e validação das peças (através do uso de algoritmos e robôs em tecnologias facilitadoras de análises de grandes conjuntos de dados); 5. a lógica da produção dos conteúdos (que impactam nas diferentes maneiras de se trabalhar no ambiente redacional voltado para a prática do jornalismo de dados. Falamos aqui do gerenciamento que envolve o desenvolvimento de peças simultaneamente).

Concluimos que houve um salto na história, no curso da vivência em sociedade que mudou suas formas de estruturas comunicativas, culturais e consequentemente, por se tratar de um ecossistema, de toda a rede. As mudanças ocorrem em diferentes etapas e hierarquias e se situar neste meio tornou-se algo extremamente relevante. Houve uma trajetória, de certa forma linear, do cinema, rádio e TV e agora há o tudo junto e misturado que envolve essa trajetória. Há um retorno, de certa forma, a oralidade que coincide com a retomada da visualidade como forma de conhecimento. As construções e apropriações ocorrem de maneira diversas não havendo regras e nem caminhos para as produções dos conteúdos (que tem, entre outros desafios, a produção em escala). Há inúmeras experimentações sejam nos modelos de negócios, nas narrativas, na interligação comunicação/tecnologia, além das tendências de outras formas de vivência como realidade aumentada, Internet das coisas, que levam ainda há outras experimentações do meio.

Neste ecossistema jornalístico, a humanização dos dados tornou-se elemento fundamental diante dos processos de mecanização. O jornalista está entre a multidão e os algoritmos exercendo o papel de um investigador, de um tradutor, de um contador de histórias. Sendo preciso entender de métricas e atravessar a barreira da alfabetização em estatística e linguagens de programação. Não defendemos que os jornalistas devam se tornar matemáticos ou programadores, mas que podem compreender alguns sistemas sem a necessidade de uma nova graduação. Sabemos do desafiador momento pelo qual o jornalismo passa financeiramente e em relação a sua credibilidade. No jornalismo de dados os desafios não são diferentes, mas é preciso, além de valorá-lo, torná-lo operacionalmente sustentável no que diz respeito aos processos de construção das peças que acontecem paralelamente e em uma lógica diferente das redações tradicionais. Acreditamos que o período experimental da visualização de dados precisa ser ultrapassado por aqueles que já a fazem, atentando para a necessidade dos gerenciamentos de projeto (como assinalado na página 176, em soluções e validações da peças).

Percebemos que é preciso dar valor e funcionalidade ao jornalismo retomando perguntas fundamentais para o exercício da profissão: o que é possível fazer hoje melhor do que o jornalismo de modelos anteriores? Que papel o jornalista desempenha melhor do que ninguém? O que os novos atores, *softwares*, fazem melhor que os jornalistas? Onde, neste processo, os jornalistas fazem a diferença? Como o jornalismo pode ser um protagonista da vida das pessoas, ser relevante ao ponto de justificar o pagamento dos seus produtos e serviços? Para nós, a visualização de dados se mostrou como um formato que tem processos próprios que dependem dos dados, das linguagens e ferramentas, locais de armazenamento e formas de resgate e encontra-se num processo não-linear enraizado na antenarrativa jornalística. O jornalismo é uma ferramenta, mas não somente técnica. As questões envolvem contextualizações. Deve estar associado ao contexto social e cultural. Não é informação primária. A apuração é a ferramenta fundamental que justifica a presença da profissão, que só terá o desenvolvimento da utilidade do conteúdo com a compreensão das relações de poder, as traduções e desvendamentos destas relações no âmbito social. Acreditamos que o jornalismo de dados e a visualização de dados são caminhos que contribuem para a credibilidade e função social da profissão.

Diante destas considerações, sem respostas conclusivas, defendemos a importância de levar estes questionamentos para dentro da universidade. Este cenário já deve fazer parte do imaginário estudantil. Quais as formas de financiamento? Como subsidiar o jornalismo além da vontade. Qual é o futuro das notícias? É o de uma lógica fragmentada? Será a recuperação e correlação de dados? É preciso encontrar o lugar social do jornalismo, onde os critérios de noticiabilidade e aceitação são outros, falamos de nichos. Por outro lado, há conteúdos que parecem não ter valor para um público consumidor, mas que apresentam um valor social relevante para ser discutido e debatido; como financiar estes conteúdos? O empreendedorismo jornalístico deve estar presente, mas antes dele a razão social, valorativa da presença do jornalismo na sociedade.

REFERÊNCIAS

AGANBEM, Giorgio. **O que é contemporâneo?** e outros ensaios. Chapecó: Editora Argos, 2009, 92p.

ALVES, Rafael M. **O Ensino de Infografia nos Cursos de Jornalismo das Universidades do Brasil e da Espanha**, 2012. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/100425/314908.pdf?sequence=1>. Acesso em: 12 de fevereiro de 2016

AMADEU, Sérgio. As políticas da sociedade informacional, propriedade imaterial e cultura digital. **Comunicação & Sociedade**. Ano 33, n. 57, p. 59-78, jan./jun. 2012. Disponível em: <https://www.metodista.br/revistas/revistas-ims/index.php/CSO/article/view/3009/2926>. Acesso em: 21 de janeiro de 2015

AMADEU, Sérgio; PIMENTEL, Thiago. Geopolítica das redes digitais, Cyberpower e a militarização da Internet. In: VIII SIMPÓSIO NACIONAL DA ABCIBER, **Oficina**, São Paulo, dezembro de 2014.

ANDERSON, Chris. BELL, Emily. SHIRKY, Clay. **Jornalismo pós-industrial – adaptação aos novos tempos**, 2012. http://www.espm.br/download/2012_revista_jornalismo/Revista_de_Jornalismo_ESP_M_5/files/assets/common/downloads/REVISTA_5.pdf

ANDERSON, Chris. BELL, Emily. SHIRKY, Clay. **Post-Industrial Journalism - Adapting to the Present**, Tow Center for Digital Journalism, Columbia Journalism School, 2013. Disponível em: <http://towcenter.org/research/post-industrial-journalism> Acesso em: 4 de fevereiro de 2015.

ARAGÃO, Alexandre. Análise das redes sociais mostra que perfis falsos influenciaram discussão na web. **Folha de São Paulo**. Disponível em: <http://www1.folha.uol.com.br/poder/2014/09/1524593-analise-das-redes-sociais-mostra-que-perfis-falsos-influenciaram-discussao-na-web.shtml>. Acesso em: 13 de dezembro de 2014.

BAIRON, Sérgio. A Comunicação nas esferas: a experiência estética e a hipermídia. **Cibercultura: Revista USP**, São Paulo. v. 86, p. pp. 18-27, jun/ago, 2010.

BAIRON, Sérgio. **O que é hipermídia**. São Paulo: Brasiliense, 2011.

BARBOSA, Suzana. Jornalismo convergente e continuum multimídia na quinta geração do jornalismo nas redes digitais. In: CANAVILHAS, João. **Notícias e Mobilidade: O Jornalismo na Era dos Dispositivos Móveis**. Livros Labcom books, 2013. p. 33-54. www.livroslabcom.ubi.pt.

BARBOSA, Suzana. **Jornalismo Digital em Base de Dados: paradigma para produtos jornalísticos digitais dinâmicos**. (tese de doutorado) UFBA/POSCOM, 2007. Disponível em: <http://tinyurl.com/nvkxk42>. Acesso em: 25 de agosto de 2013.

BARBOSA, Suzana; TORRES, Vinícius. O paradigma 'Jornalismo Digital em Base de Dados': modos de narrar, formatos e visualização para conteúdos. **Galaxia**, São Paulo. v. 25, 2013, p. 152-164, jun. Disponível em: <Disponível: <http://tinyurl.com/kmshn9d%u200E>> Acesso em: 25 agosto de 2014.

BARIANI, Bruna Barbieri. **A importância da hipermídia como experiência na produção do conhecimento**. (dissertação). ECA/USP, 2013. Disponível em: <http://www.teses.usp.br/teses/disponiveis/27/27152/tde-12092013-162438/pt-br.php>. Acesso em: 02 de julho de 2014.

BEIGUELMAN, Giselle. **Curadoria de conteúdo é o lugar do humano na internet**, 2016. Disponível em: <http://noticias.uol.com.br/opiniaio/coluna/2016/02/21/curadoria-de-conteudo-e-o-lugar-do-humano-na-internet.htm>. Acesso em: 21 de fevereiro de 2016.

BENCKLER, Yochai. **The Wealth of Networks**, 2006. Disponível em: http://www.benkler.org/Benkler_Wealth_Of_Networks.pdf. Acesso em: 5 de janeiro de 2016.

BENEVENUTO, Fabrício. AKMEIDA, Jussara M. SILVA, Altigran, S. Silva. **Explorando redes sociais online: Da coleta e análise de grandes bases de dados às aplicações**, 2011. Disponível em: <http://homepages.dcc.ufmg.br/~fabricio/download/mini-curso-sbrc11.pdf>. Acesso em: 30 de setembro de 2014.

BERNERS-LEE, Tim, HENDLER, James & LASSILA, Ora. The Semantic Web, (2011). **Scientific America**. Disponível em: http://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf. Acesso em: 20 de julho de 2014.

BERTOCCHI, Daniela. **Dos Dados aos Formatos: Um modelo teórico para o design do sistema narrativo no jornalismo digital**. (Tese de Doutorado). ECA/USP, 2013. Disponível em: <https://danielabertocchi.com/category/producao/>. Acesso em: 01 de junho de 2014.

BERTOCCHI, Daniela; SILVEIRA, Stefanie; CARMARGO, Isadora. Possibilidades narrativas em dispositivos móveis. In: CANAVILHAS, João; SATUF, Ivan. **Jornalismo para Dispositivos Móveis: produção, distribuição e consumo**. Livros Labcom books, 2015. Disponível em: <http://www.livroslabcom.ubi.pt/book/137>. Acesso em: 29 de janeiro de 2016.

BITTENCOURT, Maíra. **Grounded Theory como metodologia de pesquisa em Mídias Digitais**, 2015. Disponível em: <http://portalintercom.org.br/anais/nacional2015/resumos/R10-3479-2.pdf>. Acesso em: 17 de setembro de 2015.

BONFIM, Marcus Vinícius de Jesus. **Transparência e accountability na comunicação organizacional do setor público: impactos da Lei de Acesso à Informação**, 2014. Disponível em:

http://www.academia.edu/9981231/Transparência_e_accountability_na_comunicação_organizacional_do_setor_público_impactos_da_Lei_de_Acesso_à_Informação. Acesso em: 22 de janeiro de 2015.

BOUNEGRU, Liliana. CHAMBERS, Lucy. GRAY, Jonathan. **Manual de Jornalismo de Dados**, 2012. Disponível em: <http://datajournalismhandbook.org/pt/index.html>. Acesso em 15 de fevereiro de 2014.

BOURDIEU, Pierre. CHAMBOREDON Jean-Claude, PASSERON Jean-Claude. **Le Métier de Sociologue**, Mouton, Paris-Haia, 1973.

BOURDIEU, Pierre. O Campo Científico. In: ORTIZ, Renato. **Coleção Grandes Cientistas Sociais**. São Paulo: Ática, 1983. p.112-155.

BRADSHAW, Paul. **How to be a data journalist**, 2010. Disponível em: <http://www.guardian.co.uk/news/datablog/2010/oct/01/data-journalism-how-to-guide>. Acesso em: 10 de dezembro de 2014.

BRAGA, José Luiz. **Comunicação, disciplina indiciária**. Trabalho apresentado junto ao GT “Epistemologia da Comunicação”. Anais do XVI Encontro da Compós. Universidade Tuiuti do Paraná, Curitiba, junho 2007.

BRAMATTI, Daniel. Da fila da impressora ao Basômetro. In: DANTAS, Humberto; TOLEDO, José Roberto de; TEIXEIRA, Marco Antonio Carvalho. **Análise política & jornalismo de dados**. Rio de Janeiro: FGV, 2014. p. 219- 226.

CAIRO, Alberto. **Alberto Cairo: Data journalism needs to up its own standards**, 2014. Disponível em: <http://www.niemanlab.org/2014/07/alberto-cairo-data-journalism-needs-to-up-its-own-standards/>. Acesso em: 10 de julho de 2014.

CAIRO, Alberto. **El arte funcional: infografía e visualización de información**. Madrid: Alamut, 2012.

CAIRO, Alberto. **Infografía 2.0: visualización interactiva de información en prensa**. Madrid: Alamut, 2008.

CAIRO, Alberto. **Introdução a visualização de dados e infografia**, 2014. Disponível em: <https://www.youtube.com/watch?v=I081v2aWGpc>. Acesso em: 13 de abril de 2014.

CAIRO, Alberto. **Periodismo de precisión y visualización de datos (parte 1)**, (2011). Disponível em: <http://blogs.elpais.com/periodismo-con-futuro/2011/04/periodismo-de-precision-visualizacion-datos.html>. Acesso em: 20 de março de 2014.

CANAVILHAS, João. **O novo ecossistema mediático**, 2010. Disponível em: <http://www.bocc.ubi.pt/pag/canavilhas-joao-o-novo-ecossistema-mediatico.pdf>. Acesso em: 27 de janeiro de 2015.

CARD, Stuart; MACKINLAY, Jock; SHNEIDERMAN, Ben **Readings in information visualization: using vision to think**. San Francisco: Morgan Kaufmann Publishers, 1999.

CASE, Amber. **We are all cyborgs now**, 1999. Disponível em: https://www.ted.com/talks/amber_case_we_are_all_cyborgs_now. Acesso em: 27 de junho de 2014.

CASTELLS, Manuel. Internet e sociedade em rede, In: MORAES, Dênis (org.). **Por Uma Outra Comunicação**. Rio: Record, p. 255 a 287, 2003.

CHARAUDEAU, Patrick. **Discurso das Mídias**. Tradução de Angela M. S. Corrêa. 2. Ed., São Paulo: Contexto, 2012.

COHEN, Louis; MANION, Lawrence; MORRISON, Keith (2007). **Research Methods in Education**. Disponível em: <http://knowledgeportal.pakteachers.org/sites/knowledgeportal.pakteachers.org/files/resources/RESEARCH%20METHOD%20COHEN%20ok.pdf>. Acesso em: 20 de novembro de 2015.

COSTA, Caio T. **Um modelo de negócio para o jornalismo digital. Como os jornais devem abraçar a tecnologia, as redes sociais e os serviços de valor adicionado**. Revista de Jornalismo ESPM, n.º 9, p. 51 a 115, 2014. Disponível em: http://www.espm.br/download/2012_revista_jornalismo/Revista_de_Jornalismo_ESPM_9/. Acesso em: 08 de agosto de 2014.

COX, Donna. Metaphoric mappings: the art of visualization. In: FISHWICK, Paul. (org.) **Aesthetic Computing**. Cambridge, Mass.: MIT Press, 2006.

CRUCIANELLI, Sandra. **Herramientas Digitales para el Periodismo de Datos**, 2014. Curso massivo e aberto University of Texas at Austin em parceria com o Knight Center Journalism in the Americas. Realizado entre os dias 18 de agosto e 21 de setembro de 2014. Disponível em: <https://knightcenter.utexas.edu/es/blog/00-15690-como-usar-herramientas-del-periodismo-de-datos-inscribete-ahora-nuevo-curso-en-linea-g>.

DELEUZE, Gilles. **¿Que és un dispositivo? In: Michel Foucault, filósofo**. Barcelona: Gedisa, 1990. pp. 155 - 163. Disponível em: http://www.psi.uba.ar/academica/carrerasdegrado/psicologia/sitios_catedras/practicas_profesionales/812_clinica/cursada/archivos/deleuze_que_es_un_dispositivo.pdf

ENTINI, Carlos Eduardo. **Quem arquivará a internet para as futuras gerações?**, 2013. Disponível em: <http://acervo.estadao.com.br/noticias/acervo,quem-arquivara-a-internet-para-as-futuras-geracoes,9215,0.htm>. Acesso em: 22 de fevereiro de 2016.

ESPERANÇA, Claudio. **Visualização de dados: pesquisa e prática #CPBR8**, 2014. Palestra ministrada no Campus Party, palco Vênus, dia 07 de fevereiro de 2015. Disponível em: <http://campuse.ro/events/campus-party-brasil-2015/talk/palestra-visualizacao-de-dados-pesquisa-e-pratica-cpbr8/>. Acesso em: 07 de fevereiro de 2015.

ESTEVANIM, Mayanna. **Visualização de Dados Complexos no Jornalismo Digital: uma reflexão sobre os conceitos de dispositivo**, 2014. Disponível em: <http://www.intercom.org.br/papers/nacionais/2014/resumos/R9-0440-1.pdf>. Acesso em: 01 de setembro de 2014.

FELDMAN, Tony. **Multimedia**. London: Blueprint, 1995.

FERREIRA, Jairo. **Uma abordagem triádica dos dispositivos midiáticos**, 2006. Disponível em: <http://www.revistas.univerciencia.org/index.php/libero/article/viewFile/6112/5572>. Acesso em: 20 de novembro de 2013.

FEW, Stephen. **Abela's Folly – A Thought Confuser**, 2015. Disponível em: <https://www.perceptualedge.com/blog/?p=2080> - Abela's Folly – A Thought Confuser. Acesso em: 20 de janeiro de 2016.

FEW, Stephen. **Now you see it: simple visualization techniques for quantitative analysis**. Oakland: Analytics Press, 2009.

FLUSSER, Vilém. **O mundo codificado: por uma filosofia do design e da comunicação**. Org. Rafael Cardoso. Trad. Raquel Abi-Sâmara. São Paulo: Cosac Naify, 2007.

FOUCAULT, Michael. **Microfísica do poder**. 8. ed. Rio de Janeiro: Graal, 1979.

FREITAS, Rosiane C. PINTO, Renata M. **Programação de Alto Nível no Jornalismo de Dados**, 2015. Disponível em <http://portalintercom.org.br/anais/nacional2015/resumos/R10-3036-2.pdf>. Acesso em 12 de dezembro de 2015.

GABRIEL, Martha. Aula inaugural. Digicorp (ECA/USP). 01 de fevereiro de 2012. Notas de aula.

GABRIEL, Martha. **Marketing na era digital: conceitos, plataformas e estratégias**. São Paulo: Novatec, 2010.

GIANNELLA, Julia. R. **Dispositivo infovis: interfaces entre visualização da informação, infografia e interatividade em sites jornalísticos**, 2014. 189 f. Dissertação (Mestrado) – Escola de Comunicações e Artes, Universidade de São Paulo.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**, 2002. São Paulo: Atlas. Disponível em: <http://tinyurl.com/pcvp6r6>. Acesso em 29 de setembro de 2013.

GIL, Antônio Carlos. **Métodos e técnicas de pesquisa social**. 4. ed. São Paulo: Atlas, 1994.

GLASSER, Barney; STRAUSS, Anselm. **The Discovery of Grounded Theory: strategies for qualitative research**. Chicago/Nova York: Aldine de Gruyter, 1967.

GORDILLO, María Del Mar García; VALENCIA, Ofa Bezunartea; CRUZ, Inés Rodríguez. El Valor Agregado Periodístico, herramienta para el periodismo de calidad. In: MONPART, Josep L. Gómez; LOZANO, Juan F. Gutiérrez; SAMPÍO, Dolors Palau (Org.). **Investigar la Comunicación hoy Revisión de políticas científicas y aportaciones metodológicas**, 2013. Disponível em: <https://uvadoc.uva.es/bitstream/10324/3032/1/MetodosYTecnicasDeAnalisisYRegistroParaInvestigar.pdf>. Acesso em: 10 de outubro de 2014.

GORDON, Rich. **Data as journalism, journalism as data**, 2007. Disponível em: <http://tinyurl.com/q8a9ho6>. Acesso em: 10 de agosto de 2013.

GRANDO, Carolina Pompeo. **Os dispositivos que constituem o dispositivo**, 2012. Disponível em: http://www.observatoriodaimprensa.com.br/news/view/_ed711_os_dispositivos_que_constituem_o_dispositivo. Acesso em: 03 de dezembro de 2013.

GREY, Jonathan. BOUNEGRU, Liliana. CHAMBERS, Lucy. **Manual de Jornalismo de Dados**, 2012. Disponível em: <http://datajournalismhandbook.org/pt/>. Acesso em: 17 de novembro de 2015.

GUIMARÃES, Luciano. **A cerca do Jornalismo Visual: busca pela delimitação da investigação das imagens no jornalismo**, (2012). Disponível em: <http://www.intercom.org.br/papers/nacionais/2012/resumos/R7-2382-2.pdf>. Acesso em: 4 de setembro de 2013.

HASWANI, Mariângela Furlan. Comunicação governamental: em busca de um alicerce teórico para a realidade brasileira. In: **Revista Organicom – Dossiê Comunicação Pública e Governamental**. São Paulo: Gestcorp ECA-USP, Ano 3, n.º 4, 1º semestre, 2006.

IANNI, Octavio. Globalização: novo paradigma das ciências sociais. **Revista Estudos Avançados**, São Paulo: USP/IEA, Vol. 8, no. 21, 1994.

JANSEN, Thiago P. **Data-driven Journalism: o uso e a visualização de base de dados na produção jornalística**, 2011. Disponível em: <https://literaturaexpandida.files.wordpress.com/2011/09/monografia-final.pdf>. Acesso em: 27 de agosto de 2013.

JENKINS, Henry. **Cultura da Convergência**. São Paulo: Aleph, 2009.

JENKINS, Henry, FORD, Sam e GREEN, Joshua. **Spreadable Media: Creating Value and Meaning in a Networked Culture**. New York: NYU Press, 2013.

JENSEN, Klaus B.; JANKOWSKI, Nicholas W. **Metodologías cualitativas de investigación en comunicación de masas**. Barcelona: Bosch, cap. 2. p. 52 a 232, 1993.

JENSEN, Klaus B. Teoria e Filosofia da Comunicação. **Revista Matrizes**. Ano 2, n. 1, p. 31-49, 2008.

JOHNSON, Steven. **O mapa fantasma: como a luta de dois homens contra o cólera mudou o destino de nossas metrópoles**. Rio de Janeiro: Zahar, 2008.

KLEIN, Otavio José. **A gênese do conceito de dispositivo e sua utilização nos estudos midiáticos**, 2007. Disponível em: <http://www.ec.ubi.pt/ec/01/pdfs/klein-otavio-genese-do-conceito-de-dispositivo.pdf>. Acesso em: 11 de novembro de 2013.

KOSARA, Robert. **Stories Are Gateways Into Worlds. Eager Eyes: Visualization and Visual Communication**, 2014. Disponível em: <http://eagereyes.org/blog/2014/stories-are-gatewaysinto-worlds>>. Acesso em: 13 de abril de 2014.

KOSARA, Robert. Tableau Public. **Eager Eyes: Visualization and Visual Communication**, 2010. Disponível em: <http://eagereyes.org/blog/2010/tableau-public>>. Acesso em: 11 de fevereiro de 2014.

LANDOW, George. **Hipertexto 3.0, teoria crítica y nuevos médios en la era de la globalización**. Barcelona: Paidós Ibérica, 2009.

LEITE, Edmundo. **Datalização é o desafio na era da digitalização**, 2013. Disponível em: <http://acervo.estadao.com.br/noticias/acervo,datalizacao-e-o-desafio-na-era-da-digitalizacao,9285,0.htm>. Acesso em: 5 de dezembro de 2015.

LEITE, Edmundo. Entrevista concedida a Mayanna Estevanim, em visita ao Acervo (não gravada). Dia 17 de novembro de 2015.

LIMA, Edvaldo P. **O jornalismo impresso e a teoria geral dos sistemas: um modelo didático de abordagem**. Dissertação de Mestrado. Escola de Comunicações e Artes da Universidade de São Paulo, 1981.

LOPES, Maria Immacolata V. **Pesquisa em comunicação**. São Paulo: Loyola, 11^a ed., 2010.

LORENZ, Mirko. **Status and Outlook for a data-driven journalism. Data driven journalism: What is there to learn?**, 2010. Amsterdã: European Journalism Centre. Disponível em: <http://pt.slideshare.net/duarteromero/datadriven-journalism-what-is-there-to-learn-12056926>. Acesso em: 15 de dezembro de 2014.

MACHADO, Irene. **Comunicação de sínteses em perspectiva semiótico-evolutiva: modelização de linguagens, de formatos, do signo informático, de temporalidades**, 2010. Disponível em: http://webcache.googleusercontent.com/search?q=cache:7vpMuOsV-_kJ:revistas.unisinos.br/index.php/fronteiras/article/view/4670/1893+&cd=1&hl=en&ct=clnk&gl=br. Acesso em: 6 de fevereiro de 2016.

MACHADO, Irene. **Escola de semiótica: a experiência de Tártu-Moscou para o estudo da cultura**, São Paulo: FAPESP/ Ateliê Editorial, 2003.

MACHADO, Irene. **Gêneros e/ou formatos? Design de linguagem mediada**. Congresso Brasileiro de Ciências da Comunicação, Santos, SP, 2007.

MANOVICH, Lev. **Data stream, database, timeline** . Softwares Studies, 2012. Disponível em: <<http://lab.softwarestudies.com/2012/10/data-stream-database-timeline-new.html>> Acesso em: 10 de junho de 2014.

MANOVICH, Lev. **Data Visualisation as New Abstraction an Anti-Sublime**, 2002. Disponível em <<http://www.manovich.net>>. Acesso em: 10 de setembro de 2013.

MANOVICH, Lev. **El lenguaje de los nuevos medios de comunicación**. La imagen en la era digital, 2006. Disponível em: <https://uea.larteycomunicacion.files.wordpress.com/2013/09/manovich-el-legunaje-de-los-nuevos-medios.pdf>. Acesso em: 25 de novembro de 2013.

MANOVICH, Lev. Visualização de dados como uma nova abstração e anti-sublime. In: LEÃO, Lucia. (org.). **Derivas: cartografias do ciberespaço**. São Paulo: Annablume, 2004.

MARTÍN-BARBERO, Jesús. Uma aventura epistemológica. **Matrizes**, v. 2, n. 2, p. 143-162, 2009b. Entrevista concedida à Maria Immacolata Vassalo de Lopes.

MARTINS, Gilberto Andrade. **Estudo de caso: uma reflexão sobre a aplicabilidade em pesquisas no Brasil**, 2008. Disponível em: <http://www.revistasusp.sibi.usp.br/pdf/rco/v2n2/02.pdf>. Acesso em: 10 de agosto de 2012.

MEDITSCH, Eduardo. **O jornalismo é uma forma de conhecimento?**, 1997. Disponível em: <http://www.bocc.ubi.pt/pag/meditsch-eduardo-jornalismo-conhecimento.pdf>. Acesso em: 26 junho de 2014.

MEYER, Philip. **Precision journalism: a reporter's introduction to social Science methods**. Bloomington: Indiana University Press, 1973.

MEYER, Philip. The future of CAR: declare Cictory and Get Out. In: **When nerds and words collide: reflections on the development of Computer Assisted Reporting**. São Petesburgo: The Poyter Institute or Media Studies, 1999. Disponível em: <http://studylib.net/doc/8012136/the-future-of-car---the-university-of-north-carolina-at-c...>Acesso em: 13 de setembro de 2015.

MEYER, Philip. **The new precision journalism**. Bloomington: Indiana University Press, 1991.

MIÈGE, Bernard. Sete Considerações fundamentais sobre o espaço público contemporâneo. In: SOUSA, Mauro Wilton de. CÔRREA, Elizabeth Saad (orgs.). **Mutações no espaço público contemporâneo** – São Paulo: Paulus, 2009.

MIELNICZUK, Luciana. **Jornalismo na web: uma contribuição para o estudo do formato da notícia na escrita hipertextual**. Tese (Doutorado) – UFBA, Salvador, 2003. Disponível em: < <http://www.scribd.com/doc/12769270/Jornalismo-na-web-uma-contribuicao-para-o-estudo-doformato-da-noticia-na-escrita-hipertextual>>. Acesso em: 10 julho de 2014.

MONPART, Josep L. Gómez; LOZANO, Juan F. Gutiérrez; SAMPIO, Dolors Palau (Org.) (2013). **La calidad periodística: Teorías, investigaciones y sugerencias profesionales**. Coeditores: Universitat Autònoma de Barcelona, Universitat Jaume I, Universitat Pompeu Fabra: Aldea Global

MORIN, Edgar. O paradigma complexo. In: **Introdução ao pensamento complexo**. Porto Alegre: Sulina, p. 57 a 77, 2006.

MOUILLAUD, Maurice e PORTO, Sergio Dayrell. **O jornal: da forma ao sentido**. Brasília: Paralelo, 1997.

PALÁCIOS, Marcos. MACHADO, Elias. **Modelos de jornalismo digital**. Salvador, Calandra, 2003.

PINK, Sarah. **The Future of Antropology: engaging the senses**. London and New York: Routledge, 2006.

PIREDDU, Mario. A carne do futuro – utopia da desmaterialização. In: Pós-Humanismo: **As Relações entre o Humano e a Técnica na Época das Redes**. São Caetano do Sul: Difusão Editora, 2010.

PIRES, Nuno M. O que é a ciência de dados e porquê ela é importante para o seu negócio. **Ecommerce Brasil**, 2014. Disponível em: <http://www.ecommercebrasil.com.br/artigos/o-que-e-ciencia-de-dados-e-porque-e-la-importante-para-o-seu-negocio/>. Acesso em: 1 de janeiro de 2015.

PORTAL IMPRENSA. **Jornalistas usam novas ferramentas para filtrar boas pautas no universo Big Data**, 2015. Disponível em: http://www.portalimprensa.com.br/especialtecnologia/conteudo_tecnologia_09.asp. Acesso em 05 de novembro de 2015.

QUIGLEY, Aaron. Aesthetics of large-scale relational information visualization in practice. In FISHWICK, Paul. (org.) **Aesthetic Computing**. Cambridge, Mass.: MIT Press, 2006.

RAMALHO, Luciano. **Python: linguagem de programação**. 25 de maio de 2015. Notas de aula (disciplina de Jornalismo On-line, ECA/USP).

RAMOS, Daniela Osvald. **Formato: condição para a escrita do Jornalismo Digital em Bases de Dados. Uma contribuição da semiótica da cultura**. (Tese de Doutorado). ECA/USP, 2011. Disponível em: <http://tinyurl.com/pxgm5vy>. Acesso em: 27 de agosto de 2013.

RAMOS, Daniela; SPINELLI, Egle; ESTEVANIM, mMyanna. **Formatos digitais e qualidade no jornalismo brasileiro em base de dados**, 2014. Disponível em: <https://cobciber4.files.wordpress.com/2015/04/cobciber4atas.pdf>. Acesso e, 2 de junho de 2015.

RAMOS, Daniela Osvald. **Os textos digitais e seus sistemas modelizantes**, 2012. Disponível em: <http://tinyurl.com/o515n7f>. Acesso em: 27 de agosto de 2013

RAPAZOTE, João. **Antropologia e documentário: da escrita ao cinema**, 2007. Disponível em: http://www.doc.ubi.pt/03/artigo_joao_rapazote.pdf. Acesso em: 02 de julho de 2014.

RECUERO, Raquel. **A Conversação em Rede**. Porto Alegre: Sulina, 2012.

RECUERO, Raquel. **Métricas de Centralidade e Conversações em Redes Sociais na Internet: Desvelando Estratégias nos Debates Presidenciais de 2014**, 2014. Disponível em: <http://www.raquelrecuero.com/artigos/abciberfinal2014.pdf>. Acesso em: 10 de dezembro de 2014.

RECUERO, Raquel. **Redes Sociais Na Internet**. 2ª Ed – Porto Alegre: Sulina. (Coleção Cíbercultura), 2011.

RIBEIRO, Daniel Melo. **Visualização de dados na Internet**. Dissertação (Mestrado em Tecnologias da Inteligência e Design Digital) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2009.

ROAM, Dan (2008). **The Back of the Napkin: Solving Problems and Selling Ideas with Pictures**. New York: Penguin Books, 2008.

RODRIGUES, Adriana A. **Infografia em base de dados no jornalismo digital**, (2009). Dissertação (mestrado) - FACOM, UFBA, Salvador.

RODRIGUES, William C. **Metodologia Científica**, 2007. Disponível em: http://professor.ucg.br/SiteDocente/admin/arquivosUpload/3922/material/Willian%20Costa%20Rodrigues_metodologia_cientifica.pdf. Acesso em: 09 de agosto de 2014.

SAAD, Elizabeth. BERTOCCHI, Daniela. **A cena cibercultural do jornalismo contemporâneo: web semântica, algoritmos, aplicativos e curadoria**, 2012. Disponível em: <http://www.matrizes.usp.br/index.php/matrizes/article/view/340>. Acesso em: 5 de junho de 2013.

SAAD, Elizabeth. **Centralidade, transversalidade e resiliência: reflexões sobre as três condições da contemporaneidade digital e a epistemologia da Comunicação**, 2015. Disponível em: http://www.academia.edu/11668397/Centralidade_transversalidade_e_resiliencia_reflex%C3%B5es_sobre_as_tr%C3%AAs_condi%C3%A7%C3%B5es_da_contemporaneidade_digital_e_a_epistemologia_da_Comunica%C3%A7%C3%A3o. Acesso em: 05 de maio de 2015.

SAAD, Elizabeth. **Estratégias 2.0 para a mídia digital. Internet, informação e comunicação**. São Paulo: Editora Senac, 2003.

SAAD, Elizabeth. **Estratégias 2.0 para a mídia digital. Internet, informação e comunicação**. São Paulo: Editora Senac, 2008.

SAAD, Elizabeth. Estratégias de conteúdo para meios digitais. In: QUADROS, Claudia; CAETANO, Kati & LARANGEIRA, Álvaro (Orgs.), 2011. **Jornalismo e**

Convergência: ensino e práticas profissionais. Disponível em: <<http://www.livroslabcom.ubi.pt/book/21>> Acesso em: 22 de outubro de 2014.

SAAD, Elizabeth. **Quem Mexeu No Meu Jornalismo?** Palestra realizada pela Faculdade Cásper Líbero. Disponível em: <http://casperlibero.edu.br/assista-ao-vivo-quem-mexeu-no-meu-jornalismo/>. Acesso em: 08 de maio de 2014.

SAAD, Elizabeth. **Reflexões para uma epistemologia da Comunicação Digital**, 2008. Universidade de São Paulo. Disponível em: <http://tinyurl.com/qbrb5mu>. Acesso em: 25 de julho de 2013.

SAAD, Elizabeth. Uma reconfiguração cultural possível e viável. **Revista Matrizes**, Ano 7 – nº 1 jan./jun, 2013.

SANTAELLA, Lucia. **Comunicação ubíqua: repercussões na cultura e na educação**. São Paulo, Ed. Paulus, 2013.

SANTAELLA, Lucia. **Matrizes da linguagem do pensamento: sonora, visual, verbal**. 3. ed. São Paulo: Iluminuras: FAPESP, 2005.

SANTOS, Boaventura de Sousa. **Introdução a uma ciência pós-moderna**. Rio de Janeiro: Graal, p. 121 a 161, 1989.

SAUDO, Andréia. SALA, Lia Geovana. **Metodologia Científica** (2008). Disponível em: <http://soniaa.arq.prof.ufsc.br/arq1001metodologiacinetificaaplicada/met2008/yin.pdf>. Acesso em: 10 de agosto de 2012.

SAUSSURE, Ferdinand de. **Curso de linguística geral**. 27ª Ed. São Paulo: Cultrix, 2006.

SCOLARI, Carlos. **As interfaces não desaparecem, transformam-se**. Revista USP n. 86, Coordenadoria de Comunicação Social, Universidade de São Paulo, Julho/Julho/Agosto, 2010.

SCOLARI, Carlos. **Hipermediaciones. Elementos para una teoría de la Comunicación interactiva**, 2008. Barcelona: Gedisa, p. 29 a 118.

SERRA, Joaquim. **Manual de Teoria da Comunicação**. Labicom, Covilhã, 2007.

SERRES, Michel. **Ramos**. Rio de Janeiro, Bertrand Brasil, 2008.

SHANNON, Claude. WEAVER, Warren. **The mathematical theory of communication**. Chicago: University of Illinois Press, 1949.

SHEDROFF, Nathan. Information interaction design: a unified field theory of design. In: JACOBSON, Robert (org.). Information Design. London: MIT Press, 1999.

SLOTERDIJK, Peter, **Esferas III**, Espumas, Editorial Siruela, Barcelona, 2006.

SOUZA, Isis C. BUENO, Renato. **Análise da Evolução Temporal de Dados**

Complexos, 2013. Disponível em: http://sbbd2013.cin.ufpe.br/Proceedings/artigos/pdfs/sbbd_wtd_05.pdf. Acesso em: 20 de janeiro de 2015.

STEARNS, Josh. The Best Online Storytelling and Journalism of 2014, 2014. **Groundswell**. Disponível em: <http://stearns.wordpress.com/2015/01/05/the-best-online-storytelling-and-journalism-of-2014/>. Acesso em: 12 de janeiro de 2015.

TAKAI, Osvaldo K. ITALIANO, Isabel C. FERREIRA, João E.. **Introdução a Banco de Dados**. DCC-IME-USP, 2005. Disponível em: <http://www.ime.usp.br/~jef/apostila.pdf>. Acesso em: 5 de agosto de 2013.

TAROZZI, Massimiliano. **O que é grounded theory: metodologia de pesquisa e de teoria fundamentada nos dados**. Ed: Vozes, Petrópolis/RJ, 2011.

TEIXEIRA, Carla Cristina C.. **Criatividade, Design Thinking e Visual Thinking e sua relação com o universo da infografia e da visualização de dados**. Rio de Janeiro, 2014. 196p. Tese de Doutorado - Departamento de Artes & Design, Pontifícia Universidade Católica do Rio de Janeiro.

THIOLLENT, Michel. **Crítica metodológica, investigação social e enquete operária**. São Paulo: Polis, caps. I e II, 1980.

THOMÉ, Clarissa. **Jornalistas do ‘Estado’ recebem no Rio prêmio Esso de jornalismo**, (2015). Disponível em: <http://politica.estadao.com.br/noticias/geral,jornalistas-do-estado-recebem-no-rio-premio-esso-de-jornalismo,10000001888>. Acesso em: 14 de novembro de 2015.

TIMPONI, Raquel. **What we will, narrativas em loop e o cinema como software: um olhar mais simplista para o cinema participativo**, (2008). Disponível em: <https://www.yumpu.com/pt/document/view/12664180/what-we-will-narrativas-em-loop-eo-cinema-como-software-cencib>. Acesso em: 30 de junho de 2014.

TOLEDO, José Roberto de. **Entrevista concedida a Mayanna Estevanim, em visita a redação do blog Estadão Dados** (não gravada). Dia 17 de novembro de 2015.

TOLEDO, José Roberto de.. **Entrevista concedida a Mayanna Estevanim** (gravada). Dia 01 de março de 2016.

TOLEDO, José Roberto de. “Pirando nos dados” - prefácio. In: DANTAS, Humberto; TOLEDO, José Roberto de; TEIXEIRA, Marco Antonio Carvalho. **Análise política & jornalismo de dados**. Rio de Janeiro: FGV, 2014. p. 7 – 10.

TRÄSEL, Marcelo. **Jornalismo Guiado por Dados: relações da cultura hacker com a cultura jornalística**, 2013. Disponível em: http://compos.org.br/data/biblioteca_2065.pdf. Acesso em 30 de junho de 2014.

TRIVINHO, Eugênio R.. Obliteração estrutural da esfera pública no ciberespaço. In: SOUSA, Mauro Wilton de. CÔRREA, Elizabeth Saad (orgs.). **Mutações no espaço público contemporâneo** – São Paulo: Paulus, 2014.

TUFTE, Eduard. **Envisioning Information**. Cheshire: Graphics Press, 1990.

TUFTE, Eduard. **The Visual Display of Quantitative Information**. Cheshire: Graphics Press, 1983.

VATTIMO, Gianni. **Sociedade transparente**. Lisboa: Relógio D'Água Editores, 1992.

WEISS, Jessica. **Como o ICIJ fez uma investigação transfronteiriça, colaborativa e em grande escala**, 2015. Disponível em: <https://ijnet.org/pt-br/blog/como-o-icij-fez-uma-investiga%C3%A7%C3%A3o-transfronteiri%C3%A7a-colaborativa-e-em-grande-escala>. Acesso em: 10 de janeiro de 2016.

WIENER, Norbert. **The human use of human beings: cybernetics and society**. Cambridge, Mass.: MIT Press, 1960.

WIENER, Norbert. **Cibernética e Sociedade: O Uso Humano dos Seres Humanos**. Tradução de José Paulo Paes, São Paulo: Cultrix, 1970

WRIGHT, Alex. **The Secret History of Hypertext**. The conventional history of computing leaves out some key thinkers, 2014. Disponível em: <http://www.theatlantic.com/technology/archive/2014/05/in-search-of-the-proto-memex/371385/>. Acesso em 30 de junho de 2014.

YIN, Robert K. **Estudo de Caso: Planejamento e Métodos**. 2. ed. Porto Alegre: Bookman, 2001.