

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
ESCOLA DE COMUNICAÇÃO

CENTRO DE FILOSOFIA E CIÊNCIAS HUMANAS

JORNALISMO

***DATA-DRIVEN JOURNALISM: O USO E A
VISUALIZAÇÃO DE BASES DE DADOS NA
PRODUÇÃO JORNALÍSTICA***

THIAGO PINHEIRO JANSEN

RIO DE JANEIRO

2011

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

ESCOLA DE COMUNICAÇÃO

CENTRO DE FILOSOFIA E CIÊNCIAS HUMANAS

JORNALISMO

***DATA-DRIVEN JOURNALISM: O USO E A
VISUALIZAÇÃO DE BASES DE DADOS NA
PRODUÇÃO JORNALÍSTICA***

Monografia submetida à Banca de
Graduação como requisito para obtenção do
diploma de Comunicação Social/
Jornalismo.

THIAGO PINHEIRO JANSEN

Orientadora: Profa. Dra. Cristiane Costa

RIO DE JANEIRO

2011

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
ESCOLA DE COMUNICAÇÃO

TERMO DE APROVAÇÃO

A Comissão Examinadora, abaixo assinada, avalia a Monografia ***Data-driven journalism: o uso e a visualização de bases de dados na produção jornalística***, elaborada por Thiago Pinheiro Jansen.

Monografia examinada:

Rio de Janeiro, no dia/...../.....

Comissão Examinadora:

Orientadora: Profa. Dra. Cristiane Costa
Doutora em Comunicação pela Escola de Comunicação – UFRJ

Prof. Dr. William Braga
Doutor em Comunicação pela Escola de Comunicação – UFRJ

Prof. Me. Augusto Gazir
Avaliador convidado

RIO DE JANEIRO

2011

FICHA CATALOGRÁFICA

JANSEN, Thiago Pinheiro

Data-driven journalism: o uso e a visualização de bases de dados na produção jornalística. Rio de Janeiro, 2011.

Monografia (Graduação em Comunicação Social/ Jornalismo) – Universidade Federal do Rio de Janeiro – UFRJ, Escola de Comunicação – ECO.

JANSEN, Thiago Pinheiro. ***Data-driven journalism: o uso e a visualização de bases de dados na produção jornalística.*** Orientadora: Cristiane Costa.
Rio de Janeiro: UFRJ/ECO. Monografia em Jornalismo.

RESUMO

Esta pesquisa objetiva mostrar as diferentes etapas de evolução do uso de bases de dados nas empresas jornalísticas impressas e em seus produtos, passando pela adoção dos computadores nas redações, pela prática do Jornalismo de Precisão, das técnicas de Reportagem com o Auxílio do Computador (RAC), e culminando no desenvolvimento do *data-driven journalism*, que, sob o paradigma do Jornalismo Digital de Base de Dados (JDBD), emerge como um dos mais promissores processos de produção jornalística do século XXI. A partir da bibliografia pesquisada, do estudo de caso da cobertura do jornal britânico *The Guardian* sobre os documentos confidenciais da Guerra do Afeganistão vazados pela organização *Wikileaks*, e da entrevista com um profissional do meio jornalístico ligados ao tema, este trabalho procura ainda definir a prática do *data-driven journalism*, descrevendo suas características e apontando suas potencialidades diante das possibilidades oferecidas pela Internet e pela disseminação do movimento "open data".

AGRADECIMENTOS

A meus pais, Vera e Arlindo, os primeiros e mais importantes professores da minha vida. Obrigado por puxarem minha orelha quando foi preciso e por terem me dado espaço e tempo quando foi importante. Obrigado pela paciência, pela confiança, pelos valores e pelo amor.

Aos amigos, próximos – vocês sabem quem são – e nem tão próximos, feitos na Escola de Comunicação da UFRJ, e cuja presença foi parte fundamental de meu aprendizado nesses cinco anos de faculdade. Obrigado pelo tempo dentro e fora das salas de aula, pelas diferentes perspectivas, pelas risadas, pelos trabalhos, pelos copos de cerveja e até mesmo pelas brigas. Se termino a faculdade como uma pessoa melhor e mais rica, é por causa de vocês.

A Vivian Macedo, pela companhia e pelo carinho em meu último ano acadêmico e de vida. Com você dividi alegrias, mas também as angústias desta etapa final. Obrigado pela paciência, pelos momentos de diversão e de calma, pelos joguinhos, bejos e abraços. E obrigado por me ensinar que Niterói é quase ali e tem mais a oferecer do que aparenta.

Obrigado aos meus amigos do Colégio Pedro II, aos que permaneceram amigos e aos que a distância levou. Os conheci em um momento importante do meu crescimento e muito do que sou aprendi com vocês. Obrigado a Marcelo Machado, que, não satisfeito em ter minha companhia por sete anos de colégio, me seguiu até a faculdade – podemos ter personalidades diferentes, mas isso não nos faz menos amigos, pelo contrário. Obrigado a Vivian Andreozzi por tudo que passamos, compartilhamos e aprendemos juntos – às vezes mais próxima, às vezes mais distante, você também me ajudou a chegar até aqui.

Obrigado aos professores e demais funcionários da Escola de Comunicação da UFRJ. Com alguns aprendi muito do que levarei para minha vida profissional, com outros aprendi muito do que levarei para vida. Obrigado a professora Cristiane Costa pela orientação, observações e otimismo durante a realização deste trabalho final. Em especial, obrigado ao professor Augusto Gazir, alguém em quem se espelhar e cuja amizade levo junto com o diploma de jornalismo.

Obrigado aos que conheci no mercado de trabalho, durante minha passagem pela TV BRASIL, pelo Jornal do Brasil, pelo Instituto Brasileiro de Análises Sociais e Econômicas (IBASE) e pela Infoglobo. As experiências que estes locais me proporcionaram complementaram meu aprendizado acadêmico e me ajudaram a moldar um pouco do profissional em início de carreira que sou hoje.

SUMÁRIO

- 1. INTRODUÇÃO**
- 2. COMPUTADORES, BASES DE DADOS E JORNALISMO**
 - 2.1. A informatização das empresas e redações jornalísticas
 - 2.2. Conceituando base de dados
 - 2.3. O início da utilização de bases de dados nas empresas jornalísticas
 - 2.4. O Jornalismo de Precisão
 - 2.5. A Reportagem com o Auxílio do Computador (RAC)
 - 2.6. A Internet e a segunda fase da RAC
- 1. O PARADIGMA DO JORNALISMO DIGITAL DE BASES DE DADOS (JDBD)**
 - 3.1. O desenvolvimento do jornalismo digital
 - 3.2. Conceituando o Jornalismo Digital de Bases de Dados (JDBD)
 - 3.3. O uso de bases de dados na produção jornalística no contexto do JDBD
 - 3.4. Funções das bases de dados no contexto do JDBD
 - 3.5. Categorias do JDBD
- 1. O *DATA-DRIVEN JOURNALISM* E SUAS POTENCIALIDADES**
 - 4.1. *Data-driven journalism*
 - 4.2. Dados e o movimento do "open data"
 - 4.3. Ver para crer: a visualização de bases de dados no *data-driven journalism*
- 1. ESTUDO DE CASO: A COBERTURA DO JORNAL *THE GUARDIAN* SOBRE OS WAR LOGS DA GUERRA DO AFGANISTÃO**
 - 5.1. O que foi e o por que de sua escolha para este trabalho?
 - 5.2. Como foi a cobertura do *Guardian*?
 - 5.3. Os war logs do Afeganistão: a seleção do *Guardian*
 - 5.4. Os war logs do Afeganistão: os ataques IED em civis, forças da coalizão e tropas de Afegãos
- 1. CONSIDERAÇÕES FINAIS**
- 2. REFERÊNCIAS BIBLIOGRÁFICAS**
- 3. ANEXOS**
 - 8.1 Lista de produções de data-driven journalism
 - 8.2 Entrevista com Alberto Cairo

1. Introdução

Comunicação e tecnologia sempre caminharam e sempre caminharão interligadas. Por tabela, o mesmo se aplica ao jornalismo. Desde que o alemão Johannes Gutenberg inventou a prensa mecânica, em meados do século XV, revolucionando a comunicação humana ao permitir o registro e a disseminação de informações em larga escala, o desenvolvimento do jornalismo enquanto prática tem estado intrinsecamente ligado ao desenvolvimento tecnológico. A capacidade de jornalistas apurarem, produzirem e distribuírem informações é resultado, dentre outros fatores, das ferramentas e recursos tecnológicos de seu tempo – sejam eles meios de comunicação, ferramentas de produção e registro de conteúdo, ou sistemas de distribuição. Cerca de cinco séculos depois da prensa de Gutenberg, a tecnologia da informática se desenvolveu a tal ponto que abalou o modo de se produzir e distribuir informações – dentro e fora do jornalismo – e pavimentou o caminho para nosso atual momento: o da revolução de uma nova mídia, caracterizada pela convergência da computação e de tecnologias midiáticas, e cuja consequência está sendo a mudança de toda a cultura para formas de produção, distribuição e comunicação mediadas por computadores (MANOVICH, 2001: 19).

Esta revolução, conforme afirma o teórico Lev Manovich, é ainda mais profunda que qualquer outra pela qual já tenhamos passado porque ela afeta todos os estágios de comunicação – incluindo a aquisição, manipulação, armazenamento e distribuição de informações – e todos os tipos de mídia – textos, imagens estáticas, imagens em movimento, sons e construções espaciais. Este trabalho se foca neste momento de revolução, a partir daquela que, na conceituação de Manovich, é uma de suas formas culturais chaves: o banco de dados, ou, como se convaciona neste trabalho, a base de dados. E vai além: procura analisar as possibilidades, repercussões e os reflexos da aplicação desta forma cultural no jornalismo, mais precisamente daquele que emerge neste início de século XXI como um dos mais promissores processos de produção jornalística baseado na utilização de bases de dados: o *data-driven journalism*.

O desenvolvimento e a disseminação da Internet, no final do século XX, estão nos proporcionando um acesso sem precedentes a dados e informações que, para não nos soterrarem, precisam ser filtrados, selecionados, verificados e apresentados de uma forma comprehensível. Todo esse trabalho é papel do jornalismo, e o *data-driven journalism* tem sido apontado por profissionais do meio como uma das melhores formas

de lidar com essas funções. Mas o que exatamente é o *data-driven journalism*? Como ele se caracteriza? Quais os fatores que levaram a sua formação? Quais possibilidades ele oferece e por que tem sido alarmado como um promissor campo jornalístico? Esclarecer essas questões é um dos principais objetivos deste trabalho, que, para isso, apoia-se em uma pesquisa bibliográfica relacionada ao tema, em declarações feitas por profissionais de dentro e de fora do jornalismo, e na análise de caso de um produto específico do *data-driven journalism*.

Dessa forma, inicio este trabalho pela análise do impacto da introdução dos computadores nas redações jornalísticas, especificamente nas de jornalismo impresso, mostrando como essa tecnologia alterou a dinâmica de produção deste ambiente, o modo de se fazer notícias, a relação entre o repórter e seu texto, e as possibilidades que surgiram e passaram a ser oferecidas aos jornalistas. Tomo este momento como o ponto de partida porque é justamente nele que as bases de dados deixam de serem objetos do mundo da computação e passam também a marcar presença no jornalismo, oferecendo recursos inéditos aos profissionais da notícia e contribuindo para o desenvolvimento de novas práticas no meio, como o Jornalismo de Precisão, de Philip Meyer, e as técnicas de Reportagem com o Auxílio do Computador (RAC). Pouco a pouco, esses dois métodos passam a ocupar uma grande importância na reportagem investigativa e no jornalismo contemporâneo, servindo posteriormente como base para o desenvolvimento do *data-driven journalism*.

O próximo momento analisado neste trabalho é aquele que marca o surgimento da Internet, seu desenvolvimento e sua crescente influência nas práticas jornalísticas. Desenvolvida durante a década de 80, e popularizada nos anos 90, a rede mundial de computadores faz surgir um segmento jornalístico até então inexistente: o jornalismo digital. Em cerca de duas décadas, esta nova modalidade alcança sua terceira fase/etapa/onda/geração de desenvolvimento (PRYOR, 2002; PAVILIK, 2001; SILVA JR., 2002; MIELNICZUK, 2003), caracterizada pela expansão de iniciativas empresariais e editoriais com um foco exclusivo no online e que permitem a criação de sites jornalísticos estruturalmente desassociados do jornal impresso e completamente adaptados às características do novo meio. Mais do que a utilização de ferramentas específicas para a produção de notícias, este momento de amadurecimento do jornalismo digital é marcado pela incorporação de recursos de interatividade e personalização, pelo emprego do hipertexto na narrativa de fatos, e, principalmente, pela consolidação das bases de dados como sua estrutura chave. Como afirma Bruce

Garrison (1998), este momento representa o início da "era das bases de dados". Esta é a consolidação da fissão entre o jornalismo impresso e o jornalismo online, e o início do desenvolvimento daquilo que Suzana Barbosa (2007) denomina como o paradigma do Jornalismo Digital de Bases de Dados (JDBD), principal conteúdo do terceiro capítulo deste trabalho.

Considerado como um momento de transição entre a terceira e a quarta fase de desenvolvimento do jornalismo digital, o JDBD se estabelece como um modelo que tem as bases de dados como definidoras de sua estrutura de organização e apresentação, de acordo com funcionalidades e categoriais específicas que lhe permitem a criação, manutenção, atualização, disponibilização e circulação de produtos jornalísticos digitais dinâmicos. Conforme aponta Suzana Barbosa e outros estudiosos, é no JDBD que as possibilidades para o uso de bases de dados do jornalismo se multiplicam e se tornam explícitas, agregando funcionalidades que vão de processos produtivos à recursos administrativos de conteúdo, passando por sistemas de interação e personalização do leitor. Caracterizado por conceitos como dinamicidade, automatização, flexibilidade, inter-relacionamento/hiperlinkagem, densidade informativa, diversidade temática e visualização, o JDBD estabelece algumas das bases e condições para o desenvolvimento do *data-driven journalism* no início deste século XXI, que surge não como um produto do JDBD, mas como parte de seu escopo.

É assim que, no quarto capítulo deste trabalho, após esmiuçar nos capítulos anteriores os impactos tecnológicos, as condições e características do JDBD que culminaram no desenvolvimento do *data-driven journalism*, foco meus esforços na busca de sua definição, suas características e potencialidades. Tendo como norte as discussões levantadas por repórteres, infografistas e editores durante a mesa redonda *Data-driven journalism: What is there to learn?*, artigos dos jornalistas Rich Gordon e Paul Bradshaw, e entrevista com o infografista Alberto Cairo, identifico o *data-driven journalism* como um processo produtivo caracterizado pela convergência de diversos campos e que tem em dados "crus" a matéria-prima para a realização de sua produção, composta por análises, visualizações e narrativas. Estabeleço também em quais elementos o *data-driven journalism* se apoia para a sustentação de seus produtos: os dados, utilizados como ponto de partida, mas também em todas as etapas de produção; o processo de obtenção desses dados, por meio de técnicas de Reportagem com o Auxílio do Computador (RAC), ou por sua oferta de forma livre na Internet, na tendência do movimento "open data"; e a produção de visualizações para esses dados. Já tendo

caracterizado as técnicas de RAC no capítulo 2, parto então para análise da importância de dados "crus" e do fenômeno do "open data" – a multiplicação de movimentos, organizações e esforços pela oferta de dados livres na Internet, sejam eles governamentais ou privados – para a produção jornalística contemporânea, e traço o histórico, as potencialidades narrativas e os desafios da visualização de dados.

Por fim, para ilustrar as possibilidades sinalizadas pelo *data-driven journalism*, realizo um estudo de caso da cobertura do jornal britânico *The Guardian* sobre os relatórios secretos do exército norte-americano registrados durante a operação dos Estados Unidos no Afeganistão e que vieram a público devido à iniciativa da organização *Wikileaks*, um grupo a favor da transparência da liberdade de informações entre governos, empresas e pessoas. É importante observar, que esta análise tem um objetivo meramente expositivo e de apontar a comprovação de algumas das conclusões obtidas nos capítulos anteriores.

Em anexo, listo outros exemplos de produções de *data-driven* e disponibilizo na íntegra a entrevista feita com o infografista Alberto Cairo, também estudioso e entusiasta do *data-driven journalism*.

A importância deste trabalho repousa na crença de que compreender o *data-driven journalism*, sua formação e suas possibilidades, é compreender alguns dos caminhos que o jornalismo tomou na segunda metade do século XX – a partir adoção de computadores nas redações, da utilização de bases de dados em sua produção, da Internet e do desenvolvimento do jornalismo digital – e vislumbrar os possíveis caminhos que o jornalismo pode vir a tomar durante o século XXI.

2. Computadores, bases de dados e jornalismo

2.1 - A informatização das empresas e redações jornalísticas

A partir da década de 60, o jornalismo começou a sofrer uma impactante transformação, uma que se estende até os dias de hoje: computadores passaram a fazer parte do processo da produção de notícias nas redações e alteraram de forma profunda diversas práticas cotidianas e funções dentro das empresas jornalísticas, naquela que, para o pesquisador Anthony Smith, seria a terceira grande revolução na comunicação humana após a escrita e a invenção de Gutemberg (SMITH, 1980: 03).

Com a informatização da indústria do jornal, jornalistas e profissionais dos demais setores das empresas de notícias impressas começaram a se beneficiar das técnicas de *time-sharing*¹, passando a realizar diversas tarefas do seu cotidiano em menor tempo e com mais precisão, e os processos mecânicos de fabricação de jornais se converteram em sistemas baseados no computador (SCHEDDEN, 1998). Conforme analisado por Virgílio Gruppi Rosa, a novidade afetou a dinâmica de produção jornalística em três níveis básicos: espacial, com novas concepções para o *habitat* do jornalista; existencial, proveniente da relação entre o jornalista e seu trabalho; e profissional, referente às demissões, eliminações de etapas na produção, aumento de trabalho e responsabilidades (ROSA, 2005: 14).

Máquinas de escrever foram substituídas por computadores terminais com monitores, que passaram a ser utilizados para escrever, editar e transmitir histórias para serviços de notícias. Como consequência, as redações deixaram de serem lugares extremamente barulhentos, esfumaçados e quentes para dar lugar a espaços planejados, bem iluminados e com sistema de ar condicionado. Ainda que não tenham sumido, a quantidade de papéis empilhados diminuiu e a organização aumentou.

No espaço físico das redações a tecnologia introduziu limpeza – desapareceram as centenas de laudas amassadas no chão, sumiram as caixas de papel carbono para as cópias necessárias para a linha de produção. Até mesmo o cafezinho e o cigarro se renderam à tecnologia, uma vez que os terminais ficam

¹ O termo *time-sharing*, ou tempo compartilhado, é referente ao uso compartilhado, de maneira simultânea, por mais de uma tarefa ou usuário, dos recursos computacionais de uma máquina, como sua capacidade de processamento, seu espaço de memória, etc.

prejudicados com farelos e ambientes poluídos. Mudou também a iluminação e a temperatura do ar. Se antes do computador era inimaginável uma redação com ar condicionado e persiana nas janelas, hoje isso é rotina e já está incorporado ao dia-a-dia” (BALDESSAR, 2001: 03).

É nesse período que surge o conceito de rede: se antes a máquina de escrever era uma ferramenta isolada, com cada texto restrito àquele que o datilografava, a partir desse momento o sistema de informações se ramifica em interligações, se dispersa e descentraliza. A quantidade de dados disponíveis aumenta e a nova dinâmica de sua circulação modifica a distribuição do trabalho jornalístico em suas várias etapas, constituindo uma nova cadeia de procedimentos (ROSA, 2005: 21). Novos instrumentos de trabalho são agregados – como os terminais de computador –, cargos desaparecem – como o pauteiro –, novos são criados, alguns se transformam e, como consequência, o jornalista vê seu perfil e suas rotinas de produção – apuração, redação e edição – mudarem.

Se antes os jornalistas faziam pesquisas em extensos arquivos do jornal, matérias antigas e documentos impressos ou em microfilme, e dependiam muito de suas fontes e de sua credibilidade na apuração de suas matérias, nesse momento eles passam a ter a sua disposição ferramentas de consulta de informações diretamente em bases de dados internas do jornal, o que aumenta as possibilidades de pesquisas e reduz o tempo de procura, tornando mais práticos processos antes demorados e dispendiosos.

Ao escrever sua matéria, o jornalista deixa de datilografá-la para digitá-la, o que lhe permite um aumento no controle sobre aquilo que escreve – podendo mover e reordenar blocos de texto digitados de maneira rápida – e a redução na frustração ocasionada por erros de digitação ou na incorreta hierarquização das informações da notícia, colocando, por exemplo, o lead de uma notícia no pé do texto: em caso de equívocos, basta apagar o erro e continuar a escrever a partir dele, não havendo perda de tudo que antes fora feito até ali. Constrói-se uma nova relação entre o profissional e seu texto, caracterizada pela praticidade e agilidade.

As mudanças são percebidas não só no ambiente e na estrutura física, mas também numa nova relação com o texto. O fazer texto através do computador, com suas possibilidades de

processamento e arquivo de texto, ganha mobilidade e rapidez (...), mas é no terminal que se escondem as mais saborosas novidades para qualquer jornalista (...) para começar o usuário fica dispensado da preocupação com o fim de cada linha, o computador hifeniza (...) o computador também permite a inserção de qualquer informação, em qualquer ponto (BALDESSAR, 2001: 03).

No campo da edição, os textos encolhem, mas se ganha em recursos para facilitar o entendimento do leitor, como uma maior liberdade para a inserção de infográficos², ilustrações e fotos.

Irradiado nos Estados Unidos, o novo modelo se expande para jornais japoneses, europeus, sul-americanos e, nos anos 80, se estabelece como padrão para a indústria. No Brasil, a *Folha de S. Paulo* foi a pioneira no processo de informatização de sua redação com terminais de computador, num projeto elaborado desde 1968 e que foi concretizado somente em 1983, como parte da restruturação empresarial à qual o jornal era submetido na época e à criação de seu projeto editorial. Mais tarde, os jornais *O Globo*, *O Estado de S. Paulo*, o *Diário Catarinense*, *A Tribuna* e o *Zero Hora*, e a revista *Veja*, também instauraram processos de informatização em suas redações.

Dessa forma, o desenvolvimento tecnológico da computação e das ferramentas de telecomunicações dos últimos 50 anos acelerou a velocidade das mudanças no jornalismo, ocasionando transformações tão ou mais significativas do que aquelas desencadeadas pela prensa de Gutemberg. Atualmente, o jornalismo como indústria continua as mudanças decorrentes da década de 60, acentuadas pelo impacto da chegada da Internet na virada dos anos 80 para os 90, e marcadas pelo surgimento e desenvolvimento do jornalismo digital³, que traz consigo novos paradigmas e práticas ao meio profissional.

² Ou informações gráficas, são representações visuais de informações e dados com o objetivo de informar ao leitor de forma rápida e clara sobre um determinado assunto. Inicialmente utilizada como ferramenta por matemáticos e estatísticos, a infografia tornou-se parte importante do jornalismo a partir do século XIX.

³ Modalidade jornalística desenvolvida a partir do surgimento e disseminação da internet, durante a década de 90 do século XX.

2.2 - Conceituando base de dados

Como peça fundamental da informatização do jornalismo e das mudanças por ele sofridas na segunda metade do século XIX, e início do século XX, está a absorção e utilização da tecnologia de bases de dados e suas potenciais aplicações para a área.

De acordo com Célio C. Guimarães, uma base de dados, ou banco de dados, é uma coleção de dados ou informações relacionadas entre si, que representam aspectos do mundo real com significado próprio e que desejamos armazenar para o futuro (GUIMARÃES, 2003: 19). Sua projeção, construção e estruturação estão submetidas ao uso e conjunto de aplicações que se deseja ter sobre ela e que podem ser simples ou complexas, de tamanhos variados. Atualmente, podem conter como dados tanto informações alfanuméricas, quanto imagens, gráficos, vídeos, sons e até outras bases de dados. Bases de dados podem ser simples e pequenas, como a relação de nomes e telefones das pessoas conhecidas por um indivíduo, ou dos bens e valores de uma pessoa física; ou tão grande como o catálogo de todos os livros publicados nos Estados Unidos, ou uma base de dados de todas as fotos recolhidas ao longo dos anos pelo sistema espacial americano.

Inicialmente criadas de maneira manual, as bases de dados passaram a ser utilizadas em computadores devido ao barateamento desses equipamentos e de sua praticidade, tanto de espaço ocupado, quanto de uso, que eles oferecem. A partir de então, a complexidade das bases de dados passou a estar relacionada não necessariamente ao seu tamanho, mas ao tipo de informações que armazenam, ao modo como são manipuladas e ao uso que se faz delas.

De acordo com o crítico literário e estudioso de novas mídias Lev Manovich, as bases de dados, atualmente, são diferentes das tradicionais coleções de documentos. Elas podem ser encontradas na maioria das áreas ou dos objetos da chamada nova mídia – computadores, Internet, games, CD-ROMs, DVDs, realidade virtual, ou seja, gráficos, imagens, sons, formas, espaços e textos digitalizados – que surge por meio da convergência entre computação e tecnologia de mídia e se baseia no uso do computador como ferramenta de exibição e distribuição, mas também de produção, armazenamento e convergência. O resultado dessa nova mídia – caracterizada por conteúdos de diferentes esferas, mas que são digitalizados, ou seja, dados de computador – é a mudança de toda a cultura para formas de produção mediadas pelo computador.

Lev Manovich acredita que, nesse contexto, mais do que uma coleção de dados ou informações, as bases de dados representam atualmente uma das formas culturais simbólicas daquilo que ele denomina como "era do computador".

Similarmente, uma narrativa literária ou cinematográfica, um plano arquitetural, e uma base de dados, cada um apresenta modelos diferentes de como o mundo é. É essa percepção de base de dados como uma forma cultural de si mesmo a que eu me refiro aqui. A partir da análise do historiador de arte Ervin Panofsky sobre a perspectiva linear como uma "forma simbólica" da era moderna, nós podemos até mesmo chamar as bases de dados de uma nova forma simbólica para a era do computador (ou, como o filósofo Jean-François Lyotard a chama em seu famoso livro de 1979, *The Postmodern Condition*, uma "sociedade computadorizada"), um novo modo de estruturar as nossas próprias experiências e do mundo (MANOVICH, 2001: 219)⁴.

Assim, como forma cultural simbólica da nova mídia, as bases de dados incorporam determinadas tendências resultantes da cultura computadorizada (MANOVICH, 2001: 27-45): representação numérica, que é a capacidade de um objeto ser descrito matematicamente ou manipulado algorítmicamente; modularidade, que é a estrutura fractal de um objeto, ou seja, sua capacidade de, em conjunto, formar objetos de larga-escala, porém mantendo sua identidade unitária; automação, que é a capacidade de modificação ou criação própria, sem a intenção humana, ao menos parcialmente; variabilidade, que é a capacidade de um objeto existir em diferentes e potencialmente infinitas versões; e a transcodificação, que é a capacidade de um objeto possuir mais de um nível de representação, no caso, o da cultura humana e o da máquina.

A partir da década de 70, por meio de suas características e da tendência de computadorização das redações jornalísticas, as bases de dados passam a ser utilizadas nas empresas de mídias e a ocupar um crescente espaço na criação de produtos

⁴ No original: "Similarly, a literary or cinematic narrative, an architectural plan, and a database each present a different model of what a world is like. It is this sense of database as a cultural form of its own that I want to address here. Following art historian Ervin Panofsky's analysis of linear perspective as a "symbolic form" of the modern age, we may even call database a new symbolic form of the computer age (or, as philosopher Jean-François Lyotard called it in his famous 1979 book *The Postmodern Condition*, a "computerized society"), a new way to structure our experience of ourselves, and of the world".

noticiosos e no desenvolvimento de novas técnicas e formatos de produção jornalística Elias Machado, com base na conceituação de Lev Manovich, afirma que essa importância se consolida e se acentua durante a década de 90 e princípio do século XXI, com a explosão da Internet e o desenvolvimento do jornalismo digital: como forma cultural típica da sociedade das redes, a estrutura da base de dados se estabelece como um formato para o jornalismo digital, a partir das funções de formato para estruturação de informação, de suporte para modelos de narrativa multimídia e de memória dos conteúdos publicados (MACHADO, 2004: 01). Dessa forma, as bases de dados passam a possibilitar a exploração de recursos até então inéditos na narrativa jornalística.

2.3 - O início da utilização de bases de dados nas empresas jornalísticas

Conjuntos de dados de formatos variados ou formas culturais simbólicas, favorecidas pela chegada dos computadores às redações, as bases de dados começaram a ser empregadas na indústria do jornalismo durante a década de 70. Em um primeiro momento, como estruturas para o armazenamento de informações através da digitalização de acervo e, em seguida, como sistemas de distribuição de informações jornalísticas para o público externo e para os próprios jornalistas. Enquanto o primeiro grupo passa a ter acesso a textos e artigos produzidos pelos veículos de comunicação e outras fontes a partir de terminais de vídeo – algumas das primeiras tecnologias empregadas para entrega eletrônica de conteúdos –, o segundo grupo ganha a capacidade de pesquisar em informações de arquivo e matérias antigas de maneira mais prática e ágil, o que se reflete na possibilidade de uma contextualização mais profunda de suas notícias e reportagens.

Nesse contexto, foi marcante a criação do *The New York Times Information Bank*, em 1969, considerada a primeira base de dados de um jornal (SMITH, 1980; COLOMBO, 1991; SCHEDDEN, 1998; SQUIRRA, 1998).

A equipe do *New York Times Information Bank* e os de fora que alugam tempo de uso precisam recorrer ao uso de microfichas ou microfilmes para qualquer material mais detalhado do que o resumo... Uma coisa é desenvolver um sistema que permite aos repórteres uma checagem rápida na idade e data de nascimento de congressistas e outra coisa é ajudar empresários a aprender

detalhes das políticas ou pesquisas das atividades de mercado de companhias rivais. Mas, ao longo do necrotério de arquivos eletrônicos, os jornais poderiam desenvolver um novo tipo de serviço público, interagindo com o usuário doméstico e desempenhando um importante papel dentro da "sociedade da informação" (SMITH, 1980: 114)⁵.

Posteriormente, ainda no início da década de 70, as primeiras bases de dados e de informações em que seus usuários poderiam ter acesso por meio de um computador conectado a um modem foram criadas. Em 1973, o *Mead Data Central*, uma subsidiaria da *Mead Corporation*, uma empresa com atividades em diversos setores, entre eles tecnologia de informação, desenvolveu a *Lexix*, a primeira base de dados de texto online. Cinco anos depois, a empresa estreou a *Nexix*, a primeira base de dados online contendo conteúdo nacional de jornais como o *The New York Times*, *Washington Post*, *US. News and World Report* e da *Associated Press* (AP). Em 1972, o *Dialog*, um grande distribuidor comercial de bases de dados por rede, foi criado e, em 1979, passou a disponibilizar para seus usuários seu *National Newspaper Index*, uma base de dados bibliográfica posteriormente convertida em um serviço de texto.

Desse período, destaca-se também um novo formato de mídia e distribuição de notícias, estabelecido a partir de bases de dados e que ficou conhecido como videotexto, ou videotexto interativo. Desenvolvido pela *The British Broadcasting Corporation* (BBC) durante a década de 70, *Ceefax* foi um dos primeiros sistemas de teletexto a chegar ao mercado e consistia em um sistema de transmissão de informações, por meio do uso de um modem conectado à linha telefônica, e de exibição de conteúdo, por meio de um computador pessoal, um terminal especial ou um aparelho de televisão. Ao contrário do teletexto, que somente enviava informação em uma direção, o videotexto era capaz de trocar sinais nos dois sentidos de transmissão. Apesar dos altos investimentos que recebeu de grandes grupos de jornais nos Estados Unidos, Canadá, Japão, Alemanha, Bélgica e Espanha, e do sucesso do sistema *Minitel* na França, que ao

⁵ No original: "Personnel inside the New York Times Information Bank and outsiders renting usage on a time basis have to recourse to microfiche or microfilm for any material more detailed than the abstract... It is one thing to design a system that enables reporters to quickly check the ages and birth dates of Congressmen and another thing to help businessmen learn the details of the policies or market research activity of rival companies. But in the longer the newspaper's electronic morgue could develop as a new kind of public service, interacting with the domestic user and playing an important role within the "information society"".

final de 1986 contava com mais de 2,2 milhões de aparelhos instalados, o videotexto acabou provando-se um fracasso como nova mídia e formato de jornalismo online. Seu insucesso pode ser creditado à sua qualidade: lenta resposta ao usuário, textos em baixa resolução e pouca interação permitida (MCADAMS, 1995b *apud* GUNTER, 2003: 14). Ainda assim, o sistema foi muito importante para estabelecer as bases das novas formas de mídias que viriam a se desenvolver durante a década de 90.

2.4 - O Jornalismo de Precisão

Como explica o jornalista norte-americano Philip Meyer, a disponibilidade de computadores nas redações e a crescente utilização de bases de dados a partir da década de 70 fizeram com que um dos caminhos tomados pelo jornalismo no período fosse em direção a um método mais científico na sua produção.

O aumento da disposição de computador e fez com que grandes quantidades de informação se tornassem disponíveis 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)⁶.

Assim, Meyer pode ser considerado um pioneiro e um visionário, uma vez que, no final da década de 60, já procurava aliar o uso de computadores como ferramentas de análise de dados às suas práticas na redação. Em 1967, o jornalista, então repórter do

⁶ Disponível em: <http://www.unc.edu/~pmeyer/book/Chapter1.htm>. Acessado em: 10/08/2011. No original: "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"

Detroit Free Press, testou suas teorias a respeito da produção de um jornalismo mais científico ao utilizar-se de métodos de pesquisa das ciências sociais, em conjunto com o uso de computadores, para analisar os resultados da pesquisa de opinião pública na área metropolitana de Detroit e determinar a verdadeira causa de motins urbanos que ocorreram no verão daquele ano. Na época, havia duas teorias populares entre os editores de jornais da região sobre quais grupos estariam causando os levantes e por que: pessoas pobres e com pouca educação, como forma de expressão e obtenção de vantagens; e negros do sul da cidade, que enfrentavam dificuldades para serem assimilados na cultura do norte. Por meio da realização de pesquisas de opinião com participantes e não participantes dos motins, e do cruzamento de dados resultantes das pesquisas, o *Detroit Free Press* descobriu informações que contrariavam o até então senso comum: aqueles que haviam tido educação escolar possuíam a mesma probabilidade de participar dos levantes do que aqueles que não haviam terminado o ensino médio; e aqueles que cresceram no norte da cidade tinham mais chances de estarem participando dos levantes do que os que cresceram no sul.

Pela cobertura, intitulada *The People Beyond 12th Street*, Meyer e a equipe do *Detroit Free Press* ganharam o Prêmio Pulitzer⁷ de 1968, na categoria *Local General or Spot News*. No mesmo ano, Meyer foi intitulado "a computer reporter", pelo periódico *Newsweek*, e definiu aquilo que chamaria de *Precision Journalism*⁸, uma metodologia para a coleta de informações utilizando o método científico e as técnicas de pesquisa das ciências sociais aliados ao uso do computador.

Pouco tempo depois, em 1973, Philip Meyer lançou o primeiro livro sobre o assunto, intitulado *Precision Journalism*, onde descreveu as aplicações desta metodologia no jornalismo e, em 2001, uma reedição atualizada da obra, intitulada *The New Precision Journalism*.

O novo jornalismo de precisão é um jornalismo científico. (Na França, o termo "jornalismo de precisão" tem sido traduzido como "le journalisme scientifique.") Isso significa tratar o jornalismo como se ele fosse uma ciência, adotando método

⁷ Criado em 1917, o Prêmio Pulitzer é uma premiação norte-americana voltada para a produção jornalística impressa e online, e para a produção literária e musical. Seu nome é uma homenagem a seu criador, o editor de jornais Joseph Pulitzer.

⁸ O adjetivo "de precisão" foi empregado por Everette Dennis, em 1971, para descrever o novo tipo de jornalismo proposto por Philip Meyer, baseado em métodos científicos, e diferenciá-lo do "novo jornalismo", com enfoque literário, desenvolvido durante a década de 60.

científico, objetividade científica e ideais científicos em todo o processo de comunicação de massa. Se isso soa absurdamente pretensioso, lembre-se que a própria ciência é contida quanto aos seus alcances, suas possibilidades e possui suas próprias sanções contra a pretensão (MEYER, 1991)⁹.

Conforme descreve Meyer, o Jornalismo de Precisão encontrou nas escolas de jornalismo americanas um ambiente de fácil aceitação já que a nova técnica de se fazer reportagens demonstrava a aplicabilidade prática dos métodos de pesquisas sociais ao problema real de se conseguir reunir informações em uma sociedade cada vez mais complexa. Tal facilidade de aceitação, entretanto, não foi encontrada no ambiente profissional: ao propor uma dinâmica de produção onde os próprios jornalistas, através da análise e do cruzamento de dados, produziriam notícias, o novo método de Philip Meyer ia de encontro à passividade e inocência tradicionais ao jornalismo de até então.

A tradição estabelece que a mídia deve reportar notícias, não produzi-las. O envolvimento da mídia com pesquisas de opinião pública tem sido criticado no âmbito em que a mídia não deve fazer pesquisas, mas sim esperar pacientemente até que outras pessoas as façam e então as reportem para ela. Pesquisas de mídias também violam a regra da inocência. Um repórter deve ser uma pessoa que estabelece um olhar fresco sobre tudo, algo que ele ou ela não podem fazer se forem soterrados com muito conhecimento específico. Um jornalista, Vermont Royster disse à classe de 1967 do Nieman Fellows, deve ser um "amador profissional" (MEYER, 1991)¹⁰.

⁹ Disponível em: <http://www.unc.edu/~pmeyer/book/Chapter1.htm>. Acessado em: 10/08/2011. No original: "The new precision journalism is scientific journalism. (In France, the term "precision journalism" has been translated as "le journalisme scientifique.") It means treating journalism as if it were a science, adopting scientific method, scientific objectivity, and scientific ideals to the entire process of mass communication. If that sounds absurdly pretentious, remember that science itself is restrained about its achievements and its possibilities and has its own sanctions against pretension".

¹⁰ Disponível em: <http://www.unc.edu/~pmeyer/book/Chapter1.htm>. Acessado em: 10/08/2011. No original: "The former tradition holds that media should report news, not make news. Media involvement in public opinion polling has been criticized on the ground that the media should not do polls but should wait passively until other people do them and then report on them. Media polls also violate the innocence rule. A reporter should be a person who casts a fresh eye on everything, something that he or she cannot do if burdened by too much specialized knowledge. A journalist, Vermont Royster told the 1967 class of Nieman Fellows, should be "a professional amateur"".

Mais tarde, com a introdução nas redações do uso de planilhas eletrônicas e softwares de análise estatística de bases de dados, o *Precision Journalism*, dá forma a uma nova técnica de reportagem com o uso de computador e que será um dos pilares do jornalismo desenvolvido nos anos 80 e consolidado na década de 90: a *Computer Assisted Reporting (CAR)*, ou Reportagem com o Auxílio do Computador (RAC).

2.5 - A Reportagem com o Auxílio do Computador (RAC)

A Reportagem com o Auxílio do Computador (RAC) possui duas fases distintas de desenvolvimento. A primeira, iniciada na década de 80 e que se estende até meados da década de 90, é caracterizada pelo seu surgimento, por sua expansão e pela sistematização de suas técnicas. A segunda fase, com início na segunda metade da década de 90 e que se estende até os dias de hoje, é marcada, principalmente, pela criação e adoção de softwares mais amigáveis nos computadores, como o sistema operacional Windows 95, programas de planilhas eletrônicas, como Microsoft Excel, e pelo surgimento da Internet e a incorporação de seus recursos em suas práticas (BARBOSA, 2007: 110).

Entre as décadas de 50 e 80, foram as bases de dados que proporcionaram a criação da Reportagem com o Auxílio do Computador, a partir do *Precision Journalism* de Philip Meyer e de dois níveis de utilização de computadores nas redações jornalísticas: na busca e recuperação de informações em outros computadores; e na análise de bases de dados originais e de outras fontes de informações para os relatos jornalísticos (PAUL, 1999: 109). Ancorando principalmente reportagens investigativas, a prática da Reportagem com o Auxílio do Computador ganhou força entre a década de 70 e 80, quando suas técnicas passaram a ser definidas e sistematizadas. Neste primeiro momento, as técnicas eram utilizadas como ferramentas de obtenção de informações estatísticas, principalmente aquelas relacionadas a políticas públicas de governos, sobre temas diversos e que poderiam servir apenas como dados de contextualização ou, a partir da comparação e correlação entre informações diversas, como ferramentas de levantamento de novas informações e fatos noticiosos, até então impossíveis de serem notados de outro modo.

A Reportagem com o Auxílio do Computador não substituiu as práticas tradicionais do jornalismo, mas tem se tornado grande

parte delas. Ela também requer grande responsabilidade e vigilância. A velha máxima "Verificar, verificar, verificar", aprendida nas aulas básicas de jornalismo, se torna ainda mais crítica. "Ceticismo saudável" se torna mais importante e prática de entrevistar múltiplas fontes e de cruzar referências entre elas é ainda mais crucial (HOUSTON, 2004: 03)¹¹

Em 1985, as reportagens feitas com técnicas de RAC começaram a receber reconhecimento do *Prêmio Pulitzer*, com a vitória de Craig Flournoy e George Rodrigue, do *Dallas Morning News*, na categoria *National Reporting*, por uma série de investigações que mostraram como programas habitacionais do governo norte-americano estavam sendo utilizadas para perpetuar a segregação racial no estado do Texas. Em 1989, foi a vez do jornalista Bill Dedman, do *The Atlanta Journal-Constitution*, ganhar um *Pulitzer* com uma série de reportagens realizadas com técnicas de RAC, dessa vez na categoria *Investigative Reporting*, e que revelava políticas racistas de empréstimo de dinheiro por instituições financeiras para o pagamento de financiamentos habitacionais em Atlanta. Para Philip Meyer, as vitórias de *Prêmios Pulitzer* das reportagens de RAC mostraram como a aproximação entre as práticas jornalísticas e as técnicas científicas começavam a ser reconhecidas como algo legítimo.

O caso do Texas envolvia segregação programas públicos de habitação. A história em Atlanta documentou a discriminação federal no empréstimo para hipotecas. Os dois projetos confiaram em análise de computador de grandes bases de dados governamentais. Ambos começaram com um ponto de vista teórico: um governo comprometido com igualdade racial deveria refletir essa igualdade. Mensuração é análise mostraram que, na realidade, os resultados dos programas governamentais não refletiam igualdade. Para decidir o que mensurar, os repórteres tinham que ter uma teoria e uma hipótese

¹¹ No original: "Computer-Assisted Reporting hasn't replaced proven journalistic practices but has become a part of them. It also requires greater responsibility and vigilance. The old standard "Verify, verify, verify", which one learns in basic reporting classes, becomes even more critical. "Healthy skepticism" becomes more important, and the practice of interviewing multiple sources and cross-referencing them is more crucial".

operacional. Em outras palavras, eles deveriam pensar como cientistas (MEYER, 1991)¹²

A premiação de Bill Dedman abriu caminho para que, 17 anos depois, em 2006, o *Prêmio Pulitzer* criasse um novo conjunto de regras que permitiria aos participantes concorrerem com todo tipo de material online, incluindo bases de dados.

Difundidas nas redações durante a década de 80, as práticas de RAC passaram então a ser ensinadas em cursos e escolas de jornalismo. Criaram-se campos de estudo e organizações de pesquisa que tinham como objeto principal o desenvolvimento e a disseminação dessas técnicas. Essa expansão teve seu auge quando, em 1989, a *Investigative Reporters and Editors*¹³ (IRE) abriu o *Missouri Institute for Computer-Assisted Reporting*, na Universidade de Missouri, Estados Unidos, onde também funcionava o *National Institute for Computer-Assisted Reporting* (NICAR). Em 1992, o instituto organizou a primeira conferência sobre Reportagem com o Auxílio do Computador e, a partir daí, expandiu a iniciativa para a América Latina com a criação do *Inter-American Institute for Investigative Journalism*, em 1993.

No Brasil, as técnicas de RAC só chegariam ao país na segunda metade da década de 90, quando, em sua maioria, envolveriam análises de bases de dados públicas em planilhas eletrônicas. Acredita-se que a primeira reportagem de RAC produzida no Brasil foi publicada em 1995, no jornal *Folha de S. Paulo*, no caderno *Raio-X SP*, coordenado pelos jornalistas José Roberto de Toledo e Fernando Rodrigues, e trazia um mapeamento econômico e geográfico de distritos paulistas feito com o auxílio de planilhas eletrônicas. Em 2003, a *Associação Brasileira de Jornalismo Investigativo*¹⁴, (Abraji) organizou o seu primeiro curso sobre o assunto, aberto a jornalistas e estudantes de todo o país e ministrado por Brant Houston e Ron Nixon, da IRE. Em

¹² Disponível em: <http://www.unc.edu/~pmeyer/book/Chapter1.htm>. Acessado em: 10/08/2011. No original: "The Texas case involved segregation in public housing. The Atlanta story documented discrimination in federally insured mortgage lending. Both projects relied on computer analysis of large government databases. Both started with a theoretical framework: a government with a commitment to racial equality should produce measurable results that would reflect that equality. Measurement and analysis showed that, in reality, the results of government programs did not reflect equality.²² To decide what to measure, the reporters had to have a theory and an operational hypothesis. In other words, they were thinking like scientists".

¹³ Criado em 1975, é uma organização não lucrativa dedicada a melhorar a qualidade do jornalismo investigativo. Seu objetivo é servir como um espaço onde jornalistas de todo o mundo possam compartilhar histórias, técnicas de apuração e fontes de informações.

¹⁴ Criada em 2002 por um grupo de jornalistas brasileiros, é uma organização sem fins lucrativos que organiza congressos, seminários e oficinas com o objetivo de promover o aperfeiçoamento profissional dos jornalistas interessados no tema "investigação" (<http://www.abraji.org.br/>).

2004, a série de reportagens "Homens de Bens da ALERJ", da editoria Rio, do jornal *O Globo*, tornou-se a primeira realizada com técnicas de RAC a ganhar o *Prêmio Esso de Jornalismo*¹⁵ em sua categorial principal. Nela, os jornalistas Angelina Nunes, Alan Gripp, Carla Rocha, Dimmi Amora, Flávio Pessoa, Luiz Ernesto Magalhães e Maia Menezes realizaram uma série de levantamentos sobre os bens acumulados pelos deputados do Rio de Janeiro entre 1996 a 2001 e revelaram que 27 parlamentares tiveram aumento de mais de 100% em seus patrimônios no período.

2.6 - A Internet e a segunda fase do RAC

Para Bruce Garrison, a RAC entra em uma nova fase durante a segunda metade da década de 90, com o desenvolvimento de ferramentas de organização e análise de dados mais sofisticadas e amigáveis e, principalmente, com a difusão da Internet, que popularizou a oferta de informações por empresas e órgãos governamentais, e facilitou consulta e pesquisa em bases de terceiros entre os jornalistas (GARRISON, 1998: 265).

Concebida de forma primária em 1969, quando a *Advanced Research Projects Agency*¹⁶ (Arpa) criou a Arpanet, uma rede nacional de computadores com o objetivo de servir como garantia de comunicação emergencial caso os Estados Unidos fossem atacados por outros países durante a Guerra Fria, a Internet, tal como a conhecemos hoje, só surgiria no final da década de 80, quando Tim Berners Lee inventou a *World Wide Web* (WWW), interface baseada em hipertexto e sistemas de recursos para a Internet. Nesse meio tempo, a Agência de Comunicações e Defesa ganhou o controle da Arpanet (1975), a comunidade acadêmica se tornou a principal usuária do serviço e novas redes, como a *Because It's Time Network* (Bitnet) e a *Computer Science Network* (CSNET), começaram a surgir para suprir a demanda acadêmica. A partir da criação da WWW e de programas *browsers*¹⁷ para seu acesso, no início da década de 90, a Internet começou um movimento de acelerada expansão fora das universidades e entre os usuários domésticos, tornando-se, em alguns anos, de acordo com Barrie Gunter, uma nova mídia de massa, do ponto de vista estatístico, pelo seu volume de usuários, e forçando uma revisão conceitual do termo.

¹⁵ Criado em 1955, com o nome "Prêmio Esso de Jornalismo", é o mais tradicional e disputado programa de reconhecimento de mérito dos profissionais de imprensa do Brasil.

¹⁶ Organização do Departamento de Defesa norte-americano focada na pesquisa de informações para o serviço militar.

¹⁷ Programas responsáveis por recuperar, transmitir e apresentar informações da Internet ao usuário.

Uma distinção chave entre a Internet e outras mídias de massa tradicionais é a intercambiabilidade de produtores e receptores de conteúdo. Uma das vantagens mais conhecidas da Internet é que sua audiência pode também ser uma produtora de conteúdo. Isso representa uma separação da tradicional distinção entre produtores e receptores no contexto da mídia de massa (...) Comunicação de massa é um fenômeno relativamente novo que sempre envolveu transmissão controlada para audiências passivas que nunca tiveram nenhuma atuação ou controle sobre o conteúdo das comunicações. Com a Internet, essas características da comunicação de massa foram alteradas radicalmente. Para Strangelove (1994), a Internet é uma nova forma de comunicação de massa com uma quantidade massiva de pessoas transmitindo informações para um número massivo de receptores. A Internet deu poder a seus usuários, transformando-os em produtores e receptores de conteúdo (GUNTER, 2003: 24)¹⁸.

Durante a expansão da Internet, os sites de busca de conteúdo tiveram destaque como os principais meios de se chegar a uma informação no ambiente da online. Na busca por formas de aprimorar seus espaços gráficos e recursos para manter os usuários em suas páginas, ao invés de encaminhá-los para outras, esses sites começaram a agregar conteúdos e serviços às suas estruturas, dando forma aos grandes portais da Internet, inclusive aos de notícias, e sequência ao desenvolvimento do jornalismo digital, até então conhecido somente no formato videotexto.

O termo "portal", com o significado de "porta de entrada", começou a ser usado em 1997. Nesse ano, sites como o Yahoo!

¹⁸ No original: "One key distinction between the Internet and traditional mass media is the interchangeability of producers and receivers of content. One of the Internet's most widely touted advantages is that an audience may also be a message producer. This represents a departure from the traditional distinction between producers and receivers in the mass context (...) Mass communication is a relatively new phenomenon anyway that has always involved controlled broadcasts to passive audiences who never had any significant input or control over the content of communications. With the Internet, the characteristics of mass communication have been radically changed. For Strangelove (1994), the Internet is a new form of mass communication with massive numbers of people broadcasting information to massive numbers of receivers. The Internet has empowered its users, turning them into producers as well as receivers of content".

– criado em 1994 a partir de duas cabeças geniais, David Filo e Jerry Yang, ambos oriundos do curso de engenharia elétrica da Universidade de Stanford, na Califórnia – agregaram conteúdo e *stick applications* à página de entrada, visualizada pela maioria dos usuários pelo *browser* Netscape (FERRARI, 2003: 18).

Nesta nova fase, o jornalismo digital, ou webjornalismo, torna-se, em um primeiro momento a transposição do jornalismo impresso para a Internet e, posteriormente, um conjunto de noticiários, sites e produtos que nasceram diretamente na Internet, tendo como alvo as particularidades do público deste novo meio e como prerrogativa as especificidades desta nova mídia – a hipertextualidade, a convergência e integração de mídias (texto, áudio, vídeo, imagens, animação), a interatividade com o usuário, a personalização, a memória e a instantaneidade (BARDOEL & DEUZE, 2000; PALACIOS, 2000b; MIELNICZUK, 2003) – que exigem um modo novo e particular de se produzir notícias. O desenvolvimento do jornalismo digital será mais bem analisado no próximo capítulo deste trabalho, mas, neste momento, é pertinente mencionar como e quando surgiram os primeiros sites de notícias.

Nos Estados Unidos, os primeiros sites de empresas jornalísticas foram ao ar no início da década de 90, a partir de 1992, como simples reprodutores do conteúdo publicado no papel. *St. Louis Dispatch*, *CNN*, *Chicago Tribune* e *Charlotte Observer* foram os primeiros, seguidos de outros como *Columbus Dispatch*, *Detroit Free Press* e *The New York Times*. Experiências mais interativas e personalizadas de produtos jornalísticos na Internet só viriam em março de 1995, com o lançamento do *Personal Journal*, do *The Wall Street Journal*. Seu princípio básico era enviar textos a telas de computadores, cujo conteúdo e a formatação seriam selecionados pelo próprio assinante, conforme suas preferências. Foi nesse momento que surgiram as primeiras experiências online de *breaking news* (últimas notícias), a partir da instantaneidade do meio, integração de mídias, como imagens e áudio, e a utilização do *link* como um recurso para estruturar matérias e reportagens. Na Europa, uma das primeiras iniciativas de jornal online foi feita em 1994, com o *The Electronic Telegraph*, a versão online do *The Daily Telegraph*.

No Brasil, os sites de conteúdo não surgiram por meio de sites de busca, mas sim por iniciativa das próprias empresas jornalísticas. A razão para isso pode ser encontrada na História da comunicação brasileira, onde empresas tradicionais como o grupo Estado

(do jornal *O Estado de São Paulo*), o Grupo Folha (do jornal *Folha de S. Paulo*), as Organizações Globo (do jornal *O Globo*) e a Editora Abril, sempre se mantiveram como os maiores conglomerados de mídia do país. O primeiro site jornalístico brasileiro foi o do *Jornal do Brasil*¹⁹, criado em 1995. A ele se seguiram a versão eletrônica do jornal *O Globo*, da agência de notícias *Agência Estado*, do grupo Estado, e outros. Dentre os primeiros portais brasileiros, destacam-se os lançamentos do *Universo Online* (UOL) pelo *Grupo Folha*, do *Nutec*, e do *ZAZ*, em 1996. Ponto marcante da Internet brasileira, os portais gratuitos surgiram no início dos anos 2000, a partir da expansão da telefonia e das classes C e D, e foram um dos responsáveis pela popularização da rede no país. Nessa classe, o *IG*²⁰ foi o portal mais representativo e bem sucedido desse momento.

Para Garrison, esse período de popularização da Internet, de expansão dos tradicionais jornais para o ambiente online, de desenvolvimento do jornalismo digital e crescimento na facilidade de acesso, utilização e criação de bases de dados é o início da "era das bases dados" (GARRISON, 1998: 265), que teria como consequência uma absorção tão intensa dessas estruturas no jornalismo que, em breve, elas se tornariam intrínsecas às redações: sua distinção como uma prática jornalística especial não seria mais possível. O pensamento é semelhante ao de Suzana Barbosa, para quem o lançamento de sites fez com que as bases de dados continuassem a representar um aspecto influente no jornalismo, já que a rede "amplia o poder dos arquivos para as publicações e, principalmente, torna acessível ao público um conjunto de informações de valor precioso para auxiliar nas necessidades de pesquisa de natureza diversa".

Como forma de jornalismo mais recente, o jornalismo digital – feito no âmbito da Internet e que pressupõe a coleta, produção, publicação e disseminação de conteúdos através da web e também de dispositivos móveis como celulares, computadores de mão, iPods, etc – é a modalidade na qual o computador, as bases de dados, entre outras tecnologias, já não são considerados apenas ferramentas, mas sim, constitutivas de uma prática jornalística (BARBOSA, 2007: 142)

¹⁹ Em setembro de 2010 o Jornal do Brasil abandonou o formato impresso se tornou exclusivamente digital.

²⁰ O *Internet Group* (IG) foi adquirido em 2004 pelo grupo *Brasil Telecom* e fundido aos portais *iBest* e *BrTurbo*. Em 2010 o IG foi adquirido pela empresa de telefonia OI e continua ativo como um dos grandes portais brasileiros.

Atualmente, as técnicas de RAC já estão bastante difundidas e estabelecidas no meio jornalístico, sendo utilizadas pelos principais jornais do mundo. Os americanos *The New York Times, US. News & World Report, The Washington Post, Los Angeles Times, Charlotte Observer, Raleigh News & Observer, Miami Herald, Cleveland Plain Dealer, Kansas City Star, USA Today*, o inglês *The Guardian*, os espanhóis *EL País* e *El Mundo*, e os principais jornais brasileiros, *O Globo, O Estado de S. Paulo* e a *Folha de S. Paulo*, todos utilizam as técnicas de RAC em sua produção jornalística, muitas vezes desenvolvendo setores específicos para esta função em suas redações e editorias especiais para esses produtos. O mesmo ocorre com veículos jornalísticos televisivos e online. De fato, as técnicas de RAC estão tão difundidas que seu pioneiro, Philip Meyer, passou a considerar os termos Reportagem com o Auxílio do Computador e Jornalismo de Precisão datados durante a segunda metade da década de 90.

Depois de três décadas, chegou o tempo de afastar-se da ideia de CAR e mover-se em direção a uma ideia de ciência social. Eu não sei que título utilizar para substituir CAR. "Jornalismo de Precisão" me parece datado. Nós precisamos de algo que capture a noção de um jeito novo (...) O momento de declarar o CAR vitorioso e se mover em direção a um conceito mais novo e ambicioso chegou. Nós precisamos disso e o mundo também (MEYER, 1999: 02)²¹.

A resposta para a nova definição procurada por Philip Meyer talvez possa ser encontrada no que hoje é conhecido como *data-driven journalism*, principal foco de análise deste trabalho acadêmico e que está inserido naquilo que Suzana Barbosa denomina como Jornalismo Digital em Base de Dados (JDBD), um paradigma de transição entre a 3^a e a 4^a geração do jornalismo digital cuja análise e características serão desenvolvidas no próximo capítulo.

²¹ No original: "After three decades, the time has come to move away from the CAR fram and toward a social science frame. I don't know what label to substitute for CAR. "Precision Journalism" sounds dated. We need something that captures the notion in a fresh way (...) The time has come to declare CAR victorious and move on to a fresher , more ambitious concept. We need it, and so does the world".

3 - O Paradigma do Jornalismo Digital de Base de Dados (JDBD)

3.1 - O desenvolvimento do jornalismo digital

A partir da década de 90, a Internet se estabelece como um novo meio de massa onde o jornalismo pode desenvolver uma nova vertente de sua produção, caracterizada pelas especificidades deste novo ambiente.

Assim, nasce o jornalismo digital, ou webjornalismo, que se desenvolve durante a década de 90 e início do século XXI, paralelamente à massificação e evolução da Internet, e incorpora meios, formas e recursos até então tradicionais da produção jornalística e, a partir deles, gera novas vertentes e categorias para jornalismo.

Jornalismo digital é todo produto discursivo que constrói a realidade por meio da singularidade dos eventos, que tem como suporte as redes telemáticas ou qualquer outro tipo de tecnologia por onde se transmitam sinais numéricos e que incorpore a interação com os usuários ao longo do processo produtivo. O jornalismo digital é uma das atividades que se desenvolve no ciberespaço, que é uma visualização espacial dos sistemas mundiais de circulação e armazenamento de dados em arquivos disponibilizados nas atuais e futuras redes de comunicação (MACHADO, 2000: 19)²².

De acordo com os critérios que levam em consideração, diversos autores classificam e dividem a história do jornalismo digital em diferentes etapas de evolução. Emy Armañazanas, Javier Diaz Noci e Koldo Meso (1996) propõem uma periodização que tem como princípio a utilização de computadores em redações jornalísticas a partir da década de 50. Já John Pryor (2002) propõe uma divisão na história do jornalismo digital em ondas, cujo início se dá em 1982, com o desenvolvimento e as primeiras experimentações do serviço de videotexto.

²² No original: "En una definición sintética el periodismo digital es todo el producto discursivo que construye la realidad por medio de la singularidad de los eventos, que tiene como soporte de circulación las redes telemáticas o cualquier otro tipo de tecnología por donde se transmita señales numéricas y que incorpore la interacción con los usuarios a lo largo del proceso productivo. El periodismo digital es una de las actividades que se desarrolla en el ciberespacio, que es una visualización espacial de los sistemas mundiales de circulación y almacenamiento de datos en recorridos proporcionados por las actuales y futuras redes de comunicación".

Tendo como foco a produção de conteúdos, John Pavlik (2001) distingue três fases diferentes para o jornalismo digital, onde a primeira se estabelece pela publicação de material online produzido para outros meios, como o impresso; a segunda surge a partir da produção de conteúdo original para a Internet, com a utilização de hiperlinks e ferramentas próprias da Internet; e a terceira, a atual, se caracteriza pela produção de material noticioso específico para a Internet, desenvolvido com base nos recursos que a rede tem a oferecer para os jornalistas, durante a produção, e para os leitores, durante o consumo de conteúdo.

De maneira semelhante, José de Afonso Silva Jr. (2002), estabelece três momentos para o jornalismo digital, mas com base na relação das interfaces enquanto mediadoras do jornalismo na Internet: no primeiro momento, o Transpositivo, conteúdo e formato seguem o modelo do impresso; no segundo, o Perceptivo, ainda há o aproveitamento do conteúdo e formato do modelo impresso na produção online, porém, com uma melhor percepção da estrutura e da organização das notícias na rede; e o terceiro, mais recente, o Hipermidiático, se caracteriza pelo uso de recursos próprios da Internet na produção noticiosa – hipertextualidade, convergência de suportes, etc.

Para estabelecer o paradigma que denomina de Jornalismo Digital em Base de Dados (JDBD) como uma fase de transição entre a terceira e a quarta geração do jornalismo digital, Suzana Barbosa (2007: 146) se baseia na divisão histórica proposta por Luciana Mielniczuk (2003: 31-39). Como ressalta Mielniczuk, é importante notar que as gerações propostas por ela não são etapas necessariamente subsequentes, em que uma acaba quando a próxima outra tem início.

Antes de definir e ilustrar as etapas, é preciso salientar que essas fases não são estanques no tempo, e nem são excludentes entre si, ou seja, em um mesmo período de tempo podemos encontrar publicações jornalísticas para a *web* que se enquadram em diferentes gerações e, em uma mesma publicação, pode-se encontrar aspectos que remetem a estágios distintos. Cabe apontar que essa classificação diz respeito à trajetória do conjunto de experiências e não à evolução individual dos webjornais. Um produto recém implementado, por exemplo, pode apresentar fortes características de produtos da terceira geração, assim como um webjornal que exista desde o início dos anos 90, pode ainda manter sinais de um produto de

primeira geração, mas o contrário também pode acontecer. (MIELNICZUK, 2005: 31-32).

Assim, para Luciana Mielniczuk, o jornalismo digital teria como primeira geração a fase da Transposição, onde os produtos oferecidos seriam reproduções de partes dos grandes jornais impressos, que passavam a ocupar a Internet. Nesse momento, o conteúdo do papel era simplesmente transposto para o meio online, em intervalos de 24 horas, sem que se houvesse uma preocupação com a adequação desse produto para a Internet ou a devida exploração dos recursos oferecidos pelo meio.

Na segunda geração, devido ao desenvolvimento da estrutura técnica da Internet que segue uma tendência mundial no final dos anos 90, as características da web passam a ser exploradas de maneira mais efetiva na produção do jornalismo online, ainda que ele tenha muito do modelo impresso. Recursos como *link*, *e-mail* e hipertexto passam a ser utilizados na produção online, que adquire uma hierarquia visual própria da Internet. Este momento é denominado como fase da Metáfora devido ao fato de que, nele, o jornal impresso é uma referência para a elaboração da interface de produtos, e não somente algo a ser transposto para outro meio.

Já na terceira geração, denominada como Webjornalismo, a crescente popularização do uso da Internet e o surgimento de iniciativas empresariais e editoriais com foco exclusivo no online permitem a criação de sites jornalísticos estruturalmente desassociados do jornal impresso e completamente adaptados ao novo meio. Mais do que a utilização de ferramentas específicas da Internet, esta fase é marcada pela incorporação de recursos de interatividade e personalização, bem como pelo emprego do hipertexto na narrativa dos fatos. Aqui, estabelece-se uma fissão definitiva entre o online e o impresso e definem-se aquelas que seriam as características intrínsecas do jornalismo feito na web e explicitadas, com base em diversos pesquisadores, por Carla Schwingel.

Ao pensarmos nestas características, recorremos a pesquisadores como Armañanzas (1996), Bardoel e Deuze (2000), Bonnet (2001), Machado (2000); Mielniczuk (2003), Palacios (1999; 2002), Santos (2002) e Silva Júnior (2000) que definem a hipertextualidade, a interatividade, a personalização do conteúdo, a multimidialidade, a memória e a atualização

contínua como as potencialidades deste novo meio. De forma adicional, acrescentamos a supressão dos limites de tempo e espaço para a publicação das informações (SCHWINGEL, 2003). Pavlick (2001), por sua vez, trabalha com cinco outros aspectos que nos parecem semelhantes aos propostos acima: amplas modalidades de comunicação (texto, áudio, vídeo, gráficos, animação e até uso de vídeo em 360° graus); hipermídia (como os hiperlinks, a tecnologia digital de vinculação com a televisão e conectados a internet ou a outras redes); envolvimento da audiência (a incorporação dos usuários na produção dos conteúdos por meio da interatividade); conteúdo dinâmico (rapidez e atualização contínua, aliada à qualidade do texto); e a customização (possibilidades de personalização dos conteúdos). O autor caracteriza este tipo de jornalismo como “contextualizado” (SCHWINGEL, 2005: 02).

Cabe mencionar ainda que é neste momento que as bases de dados se estabelecem como intrínsecas às rotinas de produção jornalística online, a partir das três funções determinadas por Elias Machado (2004) para as bases de dados no jornalismo digital.

Esta terceira geração geralmente pode ser analisada pela aplicabilidade das características em produtos que são desenvolvidos no ciberespaço ou pela incorporação de rotinas diferenciadas nos processos de produção de empresas ou grupos jornalísticos que elaboram produtos veiculados no ciberespaço. O jornalismo de terceira geração, assim, é o que se utiliza de banco de dados, devido à tecnologia internet que, com as linguagens de programação dinâmicas, passaram a gerar páginas que somente existem devido às solicitações do usuário ao navegar-las, telas que puderam apresentar áreas de informações flexíveis em estruturas que possibilitam a correlação de dados, de campos informativos. (FIDALGO *apud* SCHWINGEL 2005: 02).

3.2 - Conceituando o paradigma do Jornalismo Digital em Bases de Dados (JDBD)

Suzana Machado se baseia na divisão evolutiva de jornalismo digital proposta por Luciana Mielniczuk, na definição de base de dados relacionada a concepção contemporânea de Lev Manovich (2001) para narrativa, e na ideia proposta por Elias Machado (2004, 2004, 2006) das bases de dados como um formato para o jornalismo digital, para afirmar que o jornalismo produzido na Internet vive um período de transição entre sua terceira e quarta fase evolutiva. É o que ela denomina como Jornalismo Digital em Base de Dados (JDBD).

Assim, conceituamos Jornalismo Digital em Base de Dados (JDBD) como o modelo que tem as bases de dados como definidoras da estrutura e da organização, além da apresentação dos conteúdos de natureza jornalística, de acordo com funcionalidades e categoriais específicas, que vão permitir a criação, a manutenção, a atualização, a disponibilização e a circulação de produtos jornalísticos digitais dinâmicos. Além de um paradigma na transição entre a terceira e a quarta gerações do jornalismo digital, argumentamos que este modelo tem, por outro lado, o potencial para assegurar uma metáfora particular, relacionada com os modos diferenciados para a visualização das informações (BARBOSA, 2007: 28-29).

Este momento tem na utilização da tecnologia de bases de dados para a implementação dos produtos jornalísticos uma de suas características mais marcantes. Ele também representa um paradigma de transição em que as bases de dados incorporam os recursos necessário para dinamizar o modelo de produção jornalística na Internet – por meio do desenvolvimento de novos recursos, da utilização de linguagens de programação e do desenvolvimento de sistemas de gestão de conteúdos. Por fim, mas não menos importante, ele adapta a qualidade dos produtos noticiosos às características do jornalismo feito na Internet: hipertextualidade, interatividade, multimidialidade, personalização, memória/arquivo e atualização contínua (PALACIOS, BARDOEL & DEUZE, ZAMORRA, MIENLNICZUK, SALAVERRÍA *apud* BARBOSA 2007: 152).

É importante ressaltar que a formação deste período do jornalismo digital está intrinsecamente ligada à evolução e à popularização de tecnologias computacionais e de

telecomunicação, bem como à complexificação dos processos para a implementação de produtos jornalísticos no ciberespaço e que vem se processando desde o início da segunda metade da década 90. O aumento da base tecnológica com acesso à Internet, expandido por meio da popularização das conexões de banda larga e da proliferação de plataformas móveis, como celulares smartphones, laptops, *MP3Players*, e, mais recentemente, tablets, favoreceu o consumo e a produção de produtos jornalísticos online, obrigando aos jornalistas do meio digital a explorarem cada vez mais os recursos da Internet e, como consequência, as possibilidades oferecidas pela tecnologia de bases de dados. Equipes especializadas em pesquisar, construir e administrar bases de dados são então formadas nas redações para produzir estruturas jornalísticas dinâmicas e automatizadas, assim como produtos específicos para o meio online. Jornalistas, fotógrafos e profissionais de arte e infografia passam a experimentar elementos conceituais novos para organizar suas informações e também a se preocupar com a implementação de narrativas multimídias e infografias interativas na produção, já que, neste momento, o consumidor de notícias online não quer ser apenas um leitor: ele deseja ser um usuário, que interage, manipula e contribui com o produto jornalístico. Nessa estrutura, o emprego de metadados²³ e *data mining*²⁴ para a extração de conhecimento e aplicação de novos métodos para gerar visualizações diferenciadas para os conteúdos jornalísticos se tornam essenciais.

Os avanços na tecnologia de bases de dados as tornaram a solução para compatibilizar a incorporação de recursos novos e linguagens de programação para dar forma a produtos dinâmicos e melhor elaborados, a partir do desenvolvimento de sistemas de gestão de conteúdos, visando: aperfeiçoar os sistemas de produção, assegurar maior agilidade, qualidade e descomplicar o trabalho dos jornalistas, entregando sistemas mais fáceis de operar e compatíveis com as características do jornalismo no suporte digital: hipertextualidade, interatividade, multimidialidade, personalização, memória/arquivo, atualização contínua (PALACIOS *apud* BARBOSA, 2007: 151).

²³ São informações sobre outras informações, que as complementam ou ajudam identificá-las.

²⁴ Exploração de grandes quantidades de dados à procura de padrões, regras ou sequências.

Ao consolidarem a utilização de ferramentas automatizadas e diferenciadas de publicação, e estabelecer que, para implementá-las e melhor utilizá-las, os jornalistas devem agregar características semelhantes às da figura do arquiteto de informação²⁵, as bases de dados também apontariam o caminho para a quarta fase da produção jornalística na Internet.

A partir do que foi exposto, consideramos que o Jornalismo Digital de quarta geração consolidaria a utilização de bancos de dados complexos (relacionais, voltados a objetos) através da utilização de ferramentas automatizadas e diferenciadas (sistemas para a apuração, a edição e a veiculação das informações) na produção de produtos jornalísticos. Tais ferramentas vinculariam diferentes plataformas (web, e-mail, wap) e distintos ambientes (web chats, fóruns), utilizando-se de tecnologias também diferenciadas (Twiki, PHP, Pearl, XML) de acordo com seus interesses e conveniências. O grande diferencial, sob o nosso ponto de vista, é que tais sistemas necessitam ser desenvolvidos ou por arquitetos de informação com formação em jornalismo ou por jornalistas em conjunto com profissionais de informática, e em especial na área de análise de sistemas e gerenciamento de conteúdo. Tal conjunção promoveria uma melhor integração e um produto efetivamente diferenciado. (SCHWINGEL, 2005: 10-11).

3.3 - Os usos de bases de dados no contexto do JDDB

Nesse contexto, diversos pesquisadores ampliam a dimensão conceitual das bases de dados ao propor, em seus estudos, diferentes focos de análise sobre as usabilidades dessas estruturas no jornalismo digital, seja no processo de produção, nos produtos, e, até mesmo, nos recursos que este meio oferece aos seus usuários e seus leitores.

²⁵ Termo cunhado por Rosenfeld e Morville (1998) e que se caracteriza por ser aquele profissional que possui uma visão sistêmica de um processo.

a) Sistemas de gestão e publicação de conteúdo, e gerenciadores de web sites

As utilizações mais intrínsecas de bases de dados no processo produtivo jornalístico digital talvez sejam as menos aparentes para o consumidor e leitor, porém, são de suma importância na dinâmica de produção na redação: sistemas de gestão e publicação de conteúdo, e gerenciadores de web sites jornalísticos. A criação de sistemas desses tipos específicos para o jornalismo digital passou a ser efetiva na primeira década do século XXI. Antes, as empresas jornalísticas utilizavam ferramentas padrões de publicação que não necessariamente se adequavam às necessidades de seus profissionais e que se dividiam em *Content Management System (CMS)*²⁶ ou *Portals Systems*²⁷. Integradas com o sistema de publicação, as bases de dados serviriam como peças chave para a dinâmica de produção e publicação do produto jornalístico, como explica Cláudia Schwingel.

Dentre as noções de base de dados e de banco de dados expostas por Colle (2002), pensamos que o ideal seria a utilização de um mesmo banco de dados para todas as etapas produtivas, com distintas bases de dados, já que desde a apuração até a publicação todas as informações estão em rede, em última instância. O que nos parece fundamental é a proposição de ferramentas diferenciadas para cada editoria proceder à sua apuração, com filtros e acessos a bancos de dados externos específicos. Ou seja, a necessidade é de diferentes bases de dados, pois cada função requer uma base específica. Aqui há de se fazer a distinção proposta por Fidalgo (2003) quanto aos arquivos de coleção de um jornal e os arquivos de um jornal. Para a apuração, o repórter necessitaria ter acesso a todos os arquivos do jornal, já na publicação, somente estariam disponíveis para os leitores os arquivos de coleção do jornal, ou seja, aquelas informações que foram publicadas em algum momento (SCHWINGEL, 2005: 10).

²⁶ Sistemas de administração de conteúdo.

²⁷ Sistemas de gerenciamento de portais.

b) Acesso personalizado aos produtos pelos consumidores

O autor belga Raymond Colle expõe em seus estudos as vantagens das bases de dados para os jornais digitais e para o acesso personalizado dos usuários aos sistemas documentais dos jornais, bem como o benefício da técnica de *data mining* e de KDD (Knowlage Discovery in Databases²⁸) pelos jornalistas para "descobrir os fios que unem os fatos e explicar os êxitos e fracassos em múltiplos campos" (BARBOSA *apud* COLLE, 2007: 158).

c) Estruturação e organização de informações e objetividade na produção

António Fidalgo (2003) se foca na utilização de bases de dados como ferramentas de estruturação e organização de informações em um produto digital e nas vantagens desta tecnologia para a objetividade destes veículos. Para ele, um produto estruturado sobre uma base de dados possui um conteúdo dinâmico porque este é resultado de entradas de pesquisa nas notícias cadastradas na base de dados. Quanto a vantagem para a objetividade jornalística, ela estaria nas possibilidades de arquivamento e recuperação de conteúdos permitidas pela tecnologia de bases de dados e que se refletiria na produção por meio daquilo que o autor chama de resolução semântica.

Uma primeira notícia sobre um acontecimento, que à partida surge com um determinado sentido, pode ser complementada, alterada, corrigida, à medida que outras notícias sobre o mesmo acontecimento se lhe seguem. O que, de início, tinha contornos indefinidos, deixando múltiplas hipóteses em aberto, vai ganhando sucessivamente formas cada vez mais definidas (...) À medida que forem chegando notícias subsequentes, a notícia do que ocorreu vai ganhando forma, ou seja, aumenta a sua resolução semântica (FIDALGO, 2004: 03).

²⁸ Processo de análise de dados dentro da prática de *data-mining*.

d) Formato de estruturação, suporte de narrativa multimídia e memória de conteúdos publicados

A partir do princípio de transcodificação²⁹ definido por Lev Manovich (2001), Elias Machado (2004) afirma que, como formato para o jornalismo digital, as bases de dados podem desempenhar três funções simultâneas e complementares dentro dos veículos jornalísticos digitais: a de formato para a estruturação da informação; a de suporte para modelos de narrativa multimídia; e a de memória dos conteúdos publicados.

e) Potencialidades de hipermídia e interatividade do usuário

Sob a perspectiva de Cláudia Quadros, além de melhorar a qualidade dos conteúdos jornalísticos, as bases de dados dinâmicas contribuem "na exploração das potencialidades da hipermídia e garantem a interatividade do usuário a ponto de esse assumir os papéis de receptor e emissor de uma mensagem". Dessa forma, as bases de dados seriam para o jornalismo uma espécie de "representação da democracia ao seu usuário" (QUADROS, 2005: 05-06). Um exemplo desse cenário descrito por Cláudia Quadros são as caixas de comentários dos grandes portais, que permitem que os usuários não só consumam as informações de uma notícia, mas também, a partir dela, emitam suas opiniões, corrijam dados, façam sugestões, críticas e gerem discussões com outros leitores. Todo esse processo ocorre, na maioria dos casos, em tempo real e, a partir dele, notícias ganham um novo nível de repercussão.

f) Estruturação e automatização de processos de produção no jornalismo

As bases de dados teriam também uma grande importância ao permitir a estruturação da informação e a automatização de processos de produção no jornalismo. Para o jornalista Adrian Holovaty (2005, 2006), ainda que as informações jornalísticas importantes para uma notícia não sejam essencialmente estruturadas, alguns desses dados podem ser explorados e organizados por meio da implementação da tecnologia de

²⁹ De acordo com esse conceito, todos os objetos da nova mídia podem ser traduzidos para outro formato.

bases de dados combinados a linguagens de programação. Dessa forma, ele afirma, processos de produção jornalística como a obtenção e coleta, a edição e o tratamento, e a apresentação de informações poderiam se tornar mais dinâmicos e automatizados, como ocorre, por exemplo, no site agregador de notícias *Google News*³⁰.

g) Aprimoramento da técnica de *data mining*

Por último, importante recurso para melhorar a qualidade da informação pesquisada em bases de dados e a obtenção de relações pouco aparentes entre dados distintos, a técnica de *data mining* tem sido estudada mais a fundo por Walter T. Lima Júnior (2004, 2005, 2006), que também tem analisado como essas técnicas podem ser empregadas de modos originais e diferenciados para a apresentação/visualização dos conteúdos, campo intrínseco ao *data-driven journalism*, tema principal desta dissertação acadêmica.

3.4 - Funções das bases de dados no contexto do JDBD

Com base nas possibilidades de uso das bases de dados no jornalismo digital e da observação empírica de produtos jornalísticos digitais de diferentes perfis, Suzana Barbosa estabelece 18 possíveis funcionalidades das bases de dados relacionadas aos níveis de pré-produção, produção, disponibilização/circulação, apresentação, consumo e pós-produção nos meios jornalísticos digitais (BARBOSA, 2007: 214-232):

1. Indexar e classificar as peças informativas e os objetos multimídia;
2. Integrar os processos de apuração, composição e edição dos conteúdos;
3. Conformar padrões novos para a construção das peças informativas;
4. Agilizar a produção de conteúdos, em particular os de tipo multimídia;
5. Propiciar categorias diferenciadas para a classificação externa dos conteúdos;
6. Estocar o material produzido e preservar os arquivos (memória), assegurando o processo de recuperação das informações;

³⁰ Lançado em setembro de 2002 e hoje disponível em mais de 70 edições regionais em diversos idiomas, o site é um serviço do *Google* que, por meio da implementação de algoritmos, realiza, de modo automatizado, varreduras em sites noticiosos na Internet e, com base em fatores como, frequência, local em que um material é exibido online e número de citações, determina quais materiais merecem destaque em sua página inicial.

7. Permitir usos e concepções diferenciadas para o material de arquivo;
8. Garantir a flexibilidade combinatória e o relacionamento entre os conteúdos;
9. Gerar resumos de notícias estruturados e/ou matérias de modo automatizado;
10. Armazenar anotações semânticas sobre os conteúdos inseridos;
11. 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 *data mining*;
12. Ordenar e qualificar os colaboradores e "repórteres cidadãos";
13. Orientar e apoiar o processo de apuração, coleta e contextualização dos conteúdos;
14. Regular o sistema de categorização de fontes jornalísticas;
15. Sistematizar a identificação dos profissionais da redação;
16. Cartografar o perfil dos usuários;
17. Transmitir e gerar informação para dispositivos móveis (celulares, computadores de mão, *iPods*, entre outros);
18. Implementar publicidade digital;

Essas funcionalidades estariam afinadas com a denominada terceira geração da linha evolutiva da tecnologia das bases de dados, caracterizada por possibilitar a capacidade de gestão de dados, objetos e conhecimentos, e por se adequarem, entre outras aplicações, à informação textual e à publicação digital.

3.5 - Categorias específicas do JBDB

A partir das funcionalidades listadas e do modelo de apreensão que tem no conceito de resolução semântica sua noção chave, Suzana Barbosa define sete características específica do JBDB, cujo estabelecimento, segundo a pesquisadora, "define e complementa as particularidades do JBDB" e "são essenciais para atribuir e reconhecer as rupturas, as remediações, bem como as potencialidades geradas pelo uso de base de dados no jornalismo digital" (BARBOSA, 2007: 232-244). São elas:

- **Dinamicidade** - caráter dinâmico intrínseco às bases de dados. É a partir desta categoria que as demais se desenvolvem e que o jornalismo digital

encontra novas técnicas de produção, redação, edição, apresentação de informações e a capacidade de ofertar conteúdos mais originais;

- **Automatização** - resultado direto da dinamicidade, esta categoria divide-se em automatização parcial, aplicada apenas a algumas etapas do processo de produção jornalístico; automatização procedural, quando mais etapas do processo jornalístico estão automatizadas; e automatização total, quando todos os processos funcionam sem a intervenção do jornalista;
- **Flexibilidade** - categoria que corresponde à maleabilidade da base de dados de se adaptar em diversos setores e etapas da produção jornalística;
- **Inter-relacionamento/Hiperlinkagem** - categoria ligada a capacidade de se identificar de padrões e estabelecer relações entre informações;
- **Densidade informativa** - relacionada ao conceito de resolução semântica estabelecido por António Fidalgo (FIDALGO, 2003, 2004), é, de certa forma, um parâmetro de medição do grau informativo de uma determinada notícia e que pode variar conforme novos dados são descobertos e acrescentados a ela;
- **Diversidade temática** - também relacionada ao conceito de resolução semântica, caracteriza-se pela capacidade de exploração de novos eixos temáticos conceituais pelo jornalismo digital de base de dados;
- **Visualização** - categoria que se relaciona diretamente à principal temática deste trabalho acadêmico, diz respeito aos novos e diferentes modos de representação de informações jornalísticas a partir de sua estruturação em bases de dados e de noções de metadados e metainformações.

É com base nestas características, em algumas das funcionalidades das bases de dados no jornalismo digital exploradas pelos pesquisadores mencionados, e no contexto do paradigma do JDBD definido e descrito por Suzana Barbosa, que se pode pontuar o desenvolvimento, a partir do século XXI, do *data-driven journalism*, processo de produção jornalística que é eixo temático desta monografia e que será analisado e caracterizado no próximo capítulo deste trabalho. Baseado no levantamento, análise e filtragem de informações em bases de dados abertas ao público, com o objetivo de se criar novas histórias e narrativas noticiosa com caráter multimídia, interativo e dinâmico, o *data-driven journalism* é tido por muitos pesquisadores e profissionais ligados à ciência da computação e à comunicação como um dos mais promissores campos do futuro do jornalismo digital.

"Eu acho que o *data-driven journalism* é uma das maiores de áreas de crescimento potencial no futuro do jornalismo. Muitas das discussões sobre o futuro das notícias se focam em possibilidades 'glamorosas', como jornalismo em vídeo e interatividade, mas eu frequentemente vejo o jornalismo de dados sendo ignorado.

De fato eu acredito que ele é jornalismo em sua verdadeira essência: descobrir e explorar informações que o público não tem tempo suficiente para fazê-lo, interrogando-as e fazendo-as ter sentido antes de compartilhá-las com uma audiência. Se mais jornalistas fizessem isso (ao invés de confiar em 'dados' de *press releases*) nós teríamos um público muito mais iluminado.

Minha mensagem para a próxima geração de jornalistas – ou qualquer jornalista procurando por um novo nicho ou direção – seria para aprender as habilidades e as ferramentas de interrogação de dados. Não é algo glamoroso, mas é uma habilidade que muitos jornalistas não possuem, e uma que proporciona uma vantagem no mercado" (WESTBROOK, 2010)³¹.

³¹ Adam Westbrook, em entrevista para *European Journalism Centre* (EJC), em agosto de 2010. No original: "I think data-driven journalism is one of the big potential growth areas in the future of

4 - O *Data-driven journalism* e suas potencialidades

4.1 - O *Data-driven journalism*

"A responsabilidade (de analisar dados como os liberados por governos) precisa estar com a imprensa. Jornalistas precisam ser entendedores de dados. Essas são as pessoas cujo trabalho é interpretar o que o governo está fazendo para seu povo. Antigamente, vocês costumavam conseguir boas histórias conversando com pessoas em bares, e esse ainda pode ser um método de trabalho em alguns casos. Mas agora será necessário também olhar cuidadosamente para dados e que vocês se equipem com ferramentas para analisar e selecionar o que é interessante. E mantê-los em perspectiva, ajudando as pessoas a realmente ver onde tudo se encaixa e o que está acontecendo com o país (...). *data-driven journalism* é o futuro" (LEE, 2010)³².

Em 24 de agosto de 2010, mais de 50 jornalistas investigativos, editores e acadêmicos de diversos países se encontraram em Amsterdã para participar da mesa redonda *data-driven journalism: What is there to learn?* (Jornalismo Data–driven: o que há para aprender?). Planejado pelo arquiteto de informação e jornalista multimídia

journalism. A lot of the forward-thinking discussion about the future of news focuses on the 'glamorous' possibilities, like video journalism and interactivity, but I often see data journalism being ignored.

In fact, I believe it is journalism in its truest essence: uncovering and mining through information the public do not have enough time to do themselves, interrogating it, and making sense of it before sharing it with the audience. If more journalists did this (rather than relying on 'data' from press releases) we would be a far more enlightened public.

My message to the next generation of journalists - or any journalist looking for a new niche or direction - would be to learn the skills and tools of data interrogation. It's not glamorous, but it's a skill not many journalists have, and one which will give one an edge in the market"

³² Tim Berners Lee, em 19/11/2010, em um evento de lançamento dos dados sobre os gastos do governo britânico. Disponível em: <http://www.guardian.co.uk/technology/organgrinder/2010/nov/19/berners-lee-journalism-data>. Acessado em: 12/10/2011. No original: "The responsibility needs to be with the press. Journalists need to be data-savvy. These are the people whose jobs are to interpret what government is doing to the people. So it used to be that you would get stories by chatting to people in bars, and it still might be that you'll do it that way some times. But now it's also going to be about poring over data and equipping yourself with the tools to analyze it and picking out what's interesting. And keeping it in perspective, helping people out by really seeing where it all fits together, and what's going on in the country (...)Data-driven journalism is the future".

alemão Mirko Lorenz, e organizado pelo *European Journalism Centre* (EJC)³³, em parceria com a Universidade de Amsterdã e com o auxílio financeiro do Ministério da Educação, Cultura e Ciência holandês, o seminário de um dia teve como principal objetivo discutir o que é *data-driven journalism*, suas possibilidades como um campo jornalístico emergente, além de examinar e entender as ferramentas e os processos produtivos necessários para sua realização.

Dividido em quatro sessões de palestras, com mais de 15 apresentações de profissionais de mídia, o evento abordou diferentes temas relacionados ao *data-driven journalism*, como a produção, uso, integração e visualização de dados, a possibilidade de narrativas a partir de dados, e o desenvolvimento de novos formatos para a apresentação de informações e histórias. Ao final, o conteúdo das discussões foi compilado em um documento homônimo de 78 páginas e o sucesso de sua realização proporcionou a criação, pelo *European Journalism Centre*, do site *Data Driven Journalism*³⁴, dedicado a prover informações, ferramentas, tutoriais, entrevistas e estudos de caso para aqueles jornalistas e profissionais interessados no tema. E é justamente dos resultados do seminário que surgem algumas das definições mais concisas e precisas sobre *data-driven journalism*.

"*Data journalism* é obter, reportar, realizar curadoria e publicar dados de interesse público" (STRAY, 2010)³⁵

"*Data-driven journalism* é um fluxo de trabalho que consiste dos seguintes elementos: cavar a fundo por dados e, ao procurá-los, limpá-los e estruturá-los, filtrando informações específicas, implementando maneiras de visualizá-las e transformando-as em histórias" (LORENZ, 2010)³⁶

³³ O European Journalism Centre (EJC), ou Centro de Jornalismo Europeu, é um instituto independente, internacional e não lucrativo que visa promover a prática de um jornalismo de alta qualidade a partir do treinamento de jornalistas e profissionais de mídia.

³⁴ <http://datadrivenjournalism.net/>

³⁵ Jonathan Stray, jornalista e cientista da computação, em entrevista para o *European Journalism Centre* (EJC), em agosto de 2010. Disponível em: <http://datadrivenjournalism.net/about/faq>. Acessado em: 12/10/2011. No original: "Data journalism is obtaining, reporting on, curating and publishing data in the public interest."

³⁶ Mirko Lorenz, arquiteto de informação e jornalista, em entrevista para o *European Journalism Centre* (EJC), em agosto de 2010. Disponível em: <http://datadrivenjournalism.net/about/faq>. Acessado em: 12/10/2011. No original: "Data driven journalism is a workflow that consists of the following elements: digging deep into data by scraping, cleansing and structuring it, filtering by mining for specific information, visualizing it and making a story."

"*Data journalism* é (...) a convergência de diversos campos (...) – da pesquisa investigativa e estatística, ao design e à programação" (BRADSHAW, 2010)³⁷.

"*Data-driven journalism* não deixa de ser o velho e bom jornalismo de precisão, que é aquele que usa as ferramentas conceituais – estatísticas, softwares – na ciência em geral, mas sobretudo nas ciências sociais. Ele se baseia nas análises de bases de dados, números, pesquisas, etc, para extrair de todas essas ferramentas histórias para contar aos leitores. Então não é nada novo. O que muda é a quantidade, a qualidade e as facilidades de uso das ferramentas disponíveis. Mas conceitualmente, é idêntico ao bom e velho jornalismo de precisão de Philip Meyer" (CAIRO, 2011)³⁸

O *data-driven journalism* surge então a partir da década de 2000 como uma ramificação do paradigma do Jornalismo Digital em Bases de Dados e dele apropria muitas de suas características. Seu surgimento é consequência da atual fase de desenvolvimento alcançada 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 – conforme conta Rich Gordon, professor e diretor de tecnologia digital para educação da Escola de Jornalismo de Medill, na Universidade de Northwestern.

Nos anos 90, algumas das pessoas responsáveis por lançar sites de notícias (incluindo eu, como o primeiro diretor de novas mídias no *The Miami Herald*) vieram de experiências com reportagens assistidas por computador. Nós vimos o potencial das bases de dados online, mas descobrimos que uma variedade de fatores – incluindo tecnologia complexa e cara, e controle corporativo ao acesso de servidores Web – afastaram esses tipos

³⁷ Paul Bradshaw, escritor de jornalismo de dados e escritor, em entrevista para o *European Journalism Centre* (EJC), em agosto de 2010. Disponível em: <http://datadrivenjournalism.net/about/faq>. Acessado em: 12/10/2011. No original: "Data journalism is [...] the convergence of a number of fields [...] - from investigative research and statistics to design and programming."

³⁸ Alberto Cairo, diretor de infografia e multimídia da Revista Época, em entrevista realizada pelo autor em 26/11/2011.

de projetos para banho-maria. Agora que a tecnologia se tornou mais acessível, é excitante ver tantas organizações de notícias se envolvendo em publicações de bases de dados (GORDON, 2007)³⁹.

Como menciona Gordon, apesar de ter se intensificado nos últimos anos, o uso de bases de dados na produção jornalística não é algo absolutamente novo e nem recente – vem sendo feito desde a década de 70 por repórteres investigativos especializados nas práticas de Reportagem Auxiliada por Computador (RAC). Então, qual seria a diferença entre esta forma de utilização de bases de dados e aquela feita atualmente por meio do que é denominado *data-driven journalism*? De acordo com a apresentação realizada por Mirko Lorenz no seminário em Amsterdã, enquanto a prática do RAC é primariamente uma técnica, o *data-driven journalism* seria um processo de produção em que dados serviriam como base para análises, visualizações e, "mais importante", narrativas.

Então, *data-driven journalism* pode ser visto como um processo de refinamento, onde dados crus são transformados em algo com significado. Como resultado, o valor para o público cresce, especialmente quando fatos complexos são moldados em histórias claras, que as pessoas podem facilmente entender e lembrar (LORENZ, 2010: 12)⁴⁰.

Para o jornalista e escritor Paul Bradshaw, especializado em jornalismo de dados, este "processo de refinamento" representa a convergência de diversos campos significativos em um processo produtivo cada vez mais importante para as práticas jornalísticas digitais.

³⁹ Disponível em: <http://getsmart.readership.org/2007/11/data-as-journalism-journalism-as-data.html>. Acessado em: 16/10/2011. No original: "Back in the mid-1990s, quite a few of the people given responsibility for launching online news sites (including me, as the first new media director at The Miami Herald) came from computer-assisted reporting backgrounds. We saw the potential for online databases but found that a variety of factors - including complex, expensive technology and corporate control of access to Web servers - pushed these kinds of projects to the back burner. Now that the technology has gotten more accessible, it's exciting to see so many news organizations get involved in database publishing"

⁴⁰ No original: "So data-driven journalism can be viewed as a process of refinement, where raw data is transformed into something meaningful. As a result the value to the public grows, especially when complex facts are boiled down into a clear story that people can easily understand"

Jornalismo de dados é imenso. Eu não quero dizer 'imenso' como algo que está na moda – apesar de ter se tornado isso nos últimos meses – mas 'imenso' como em 'incompreensivelmente enorme'. Ele representa a convergência de diversos campos que são significativos por si só – da pesquisa investigativa e estatística a design e programação. A ideia de combinar essas habilidades para contar histórias importantes é poderosa – mas também intimidadora (BRADSHAW, 2010)⁴¹.

Essa convergência de campos se reflete nas próprias etapas de produção do *data-driven journalism*. Em sua apresentação no seminário de Amsterdã, Mirko Lorenz mostrou um gráfico onde descreveu o *data-driven journalism* como um processo que tem os dados como ponto inicial, passa por etapas de filtragem e visualização, e culmina na concepção de uma história. De maneira semelhante, Paul Bradshaw descreve o processo de produção do jornalismo de dados em quatro etapas: encontrar os dados, por meio de conhecimentos, contatos e da utilização de computadores e de habilidades técnicas específicas; interrogar os dados, para conquistar um bom entendimento de seus jargões e uma visão ampla do contexto a que se aplicam; mesclar e dar visualização aos dados, por meio de infográficos; e mesclar e apresentar os dados, por meio de ferramentas próprias ou abertas.

Ao proporcionar o casamento efetivo de reportagem tradicional e narrativa com o desenvolvimento das capacidades das bases de dados, a partir da convergência de diferentes áreas, o *data-driven journalism* representa uma nova forma de jornalismo digital. Ainda, de acordo com a hierarquização proposta por Rich Gordon para as diferentes produções do jornalismo de base de dados, o *data-driven journalism*, por suas características, ocuparia o nível mais elevado.

Mas está começando a ficar claro que existe uma grande variedade de possibilidades para a publicação de bases de dados. Alguns projetos são claramente mais complexos e potencialmente mais recompensadores – para as organizações

⁴¹ Disponível em: <http://www.guardian.co.uk/news/datablog/2010/oct/01/data-journalism-how-to-guide>. Acessado em: 10/10/2011. No original: "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?"

de notícias e sua audiência online – que outros. Pode ser útil pensar sobre uma hierarquia de publicação de bases de dados.

Na parte de baixo estão os projetos do tipo mais simples nos quais as organizações de notícias não fazem muito mais do que tornar os dados disponíveis. No topo, estão as aplicações mais ambiciosas, em que as organizações adicionam valor por meio do desenvolvimento de uma interface inteligente, análise jornalística, apresentação criativa ou conexões de narrativa (GORDON, 2007)⁴².

De acordo com Gordon, a hierarquia para os produtos do jornalismo de base de dados se organiza em cinco leveis, definidos conforme o grau de complexidade do uso que fazem das bases de dados. É possível notar uma semelhança entre os leveis definidos por Gordon e os processos de produção do *data-driven journalism* mencionados por Mirko Lorenz e Paul Bradshaw. Quanto mais longe nos processos de produção um determinado veículo vai, mais ele se aproxima do nível mais alto na hierarquia descrita por Gordon. De acordo com o jornalista, esses leveis seriam:

- **Level 1: Entrega de dados** – Aqui as organizações de notícias obtêm os dados e os tornam disponíveis de uma forma navegável. Não há reportagem adicional ou funcionalidade para o usuário online.
- **Level 2: Pesquisa de dados** – Este é o modo mais comum de disponibilizar dados. Usuários podem esperar encontrar informações relevantes ao digitar entradas em uma caixa de texto.
- **Level 3: Exploração de dados** – Casos em que é disponibilizado uma caixa de busca na página da base de dados, mas a própria página permite

⁴² Disponível em: <http://getsmart.readership.org/2007/11/data-as-journalism-journalism-as-data.html>. Acessado em: 16/10/2011. No original: "But it's also becoming clear that there's a wide range of possibilities for database publishing. And that some projects are clearly both more complex and potentially more rewarding - for news organizations and their online audience - than others. It might be useful to think about a hierarchy of database publishing. At the low end are the simplest kinds of projects in which the news organization doesn't do much beyond making the data available. At the high end are the most ambitious applications, in which the news organization adds value through smart interface development, journalistic analysis, creativity in presentation or connections to storytelling" .

uma exploração fácil dos dados de um jeito que a maioria das bases de dados online não permite.

- **Level 4: Visualização de dados** – Para muitas bases de dados, a coisa mais valiosa que uma organização de notícias pode fazer é oferecer um modo de as pessoas visualizarem o que os dados estão mostrando. A abordagem mais óbvia envolve mapeamento, ao menos para bases de dados que possuem elementos geográficos. Mas as possibilidades para a visualização de dados vão além do mapeamento.
- **Level 5: Experiência em dados e narrativas** – Quando as organizações de notícias conseguem efetivamente casar reportagem tradicional e narrativa com as capacidades do desenvolvimento de bases de dados, verdadeiras novas formas de jornalismo podem emergir.

É importante notar que, uma vez realizados em conjunto, esses processos descritos geram produtos que embarcam, em maior ou menor grau, as sete características descritas por Suzana Barbosa em seu trabalho sobre o Jornalismo Digital em Base de Dados (JDBD), e mencionadas no capítulo anterior deste trabalho: dinamicidade, automatização, flexibilidade, inter-relacionamento/hiperlinkagem, densidade informativa, diversidade temática e visualização. Esse fato evidencia o *data-driven journalism* como uma modalidade jornalística inerente ao JDBD e que aponta para a quarta fase do jornalismo digital.

Por meio das definições apresentadas a respeito de *data driven journalism*, ou ainda pela descrição de suas etapas produtivas realizadas pelos autores mencionados, é possível denotar que pelo menos três elementos são chaves nas produções baseadas em *data-driven journalism*: os dados propriamente, utilizados como ponto de partida, mas também em todas as etapas de produção; o processo de obtenção desses dados, por meio de técnicas de Reportagem com o Auxílio do Computador, ou por sua oferta de forma livre na Internet; e a produção de modos de visualização desses dados. A seguir, explorarei mais a fundo o primeiro – com base na tendência de liberdade de dados ("open data") – e o terceiro item, já que o surgimento e evolução das técnicas de Reportagem com o Auxílio do Computador já foram abordadas anteriormente.

4.2 - Os dados e o movimento "open data"

"Então eu tenho trabalhado como jornalista de dados por mais ou menos um ano e sempre escuto uma frase que é a seguinte: "dados são o novo petróleo". E dados é o tipo de recurso onipresente que podemos formatar para prover novas inovações e novas perspectivas, e está ao redor, e pode ser extraído facilmente. Essa não é nenhuma grande metáfora nesses tempos, especialmente se você vive perto do Golfo do México. Mas eu acho que adaptaria um pouco essa metáfora e diria que dados são o novo solo. Porque para mim parece uma fértil e criativa mídia. Sabe, ao longo dos anos, online, nós plantamos uma quantidade gigantesca de informações e dados, e irrigamos com redes e conectividade, e isso foi feito por empregados e governos sem pagamento (...) Mas (dados) são um meio muito fértil e parece que visualizações, infográficos, visualizações de dados eles parecem com flores desabrochando desse meio. Mas se você olhar diretamente é só um monte de números e fatos desconexos. Mas se você começar a trabalhar com eles e manipulá-los de um certo modo coisas interessantes poderão aparecer, e padrões diferentes poderão ser revelados" (MCCANDLESS, 2010)⁴³.

Em setembro de 2006, o desenvolvedor Web e jornalista Adrian Holovaty publicou em seu site pessoal o artigo *A fundamental way newspaper sites need to change*. Nele, Holovaty propõe que o jornalismo precisa parar de ser uma visão de mundo centrada em histórias, algo condicionado por décadas como um estilo

⁴³ David McCandless, jornalista, como parte da apresentação *The beauty of data visualization*, realizada em julho de 2010, TEDxGlobal 2010. Disponível em: http://www.ted.com/talks/lang/eng/david_mccandless_the_beauty_of_data_visualization.html No original: "So, I've been working as a data journalist for about a year, and I keep hearing the a phrase all the time, which is this: 'Data is the new oil'. And data is kind of ubiquitous resource, that we can shape to provide new innovations and new insights, and it's all around us, and it can be mined very easily. And it's not a particularly great metaphor in these times, especially if you live around the Gulf of Mexico, but I would, perhaps, adapt this metaphor slightly, and I would say that data is the new soil. Because for me, it feels like a fertile, creative medium. You know, over the years, online, we've laid down a huge amount of information and data, and we irrigate it with networks and connectivity, and it's been worked and tilled by unpaid workers and governments (...) But it's (data) a really fertile medium, and it feels like visualizations, infographics, data visualizations, they feel like flowers blooming from this medium. But if you look at it directly, it's just a lot of numbers and disconnected facts. But if you start working with it and playing with it in a certain way, interesting things can appear and different patterns can be revealed".

estabelecido do jornalismo. Conforme ele aponta, jornalistas sempre coletaram no seu trabalho diário aquilo que ele define como informação estruturada – um tipo de informação que pode ser destrinchada, de forma automata, por computadores. Acontece que, para o autor, chegamos a um momento da contemporaneidade em que temos recursos para apresentar esse tipo de informação de maneira mais eficaz e dinâmica, um momento em que as histórias jornalísticas simplesmente já não são mais suficientes: é preciso também oferecer esse tipo de informação aos leitores de um modo que possam ser melhor compreendidas por meio de outras formas que não a textual, já que se perdem e se diluem dentro de histórias.

Por exemplo, digamos que um jornal escreveu uma história sobre um incêndio local. Poder ler essa história em um celular é ótimo. Viva a tecnologia! Mas o que eu realmente quero é poder explorar os fatos crus dessa história, um por um, com camadas de distribuição, e uma infra-estrutura para comparar os detalhes do incêndio – data, tempo, local, vítimas, número do quartel que atendeu ao chamado, distância do quartel, nomes e anos de experiência dos bombeiros que trabalharam na cena, o tempo que levou para os bombeiros chegarem – com detalhes de incêndios passados. E também com incêndios futuros, quando eles ocorrerem.

Isso é o que eu quero dizer com dados estruturados: informação com atributos que são consistentes em um domínio. Todo incêndio tem os seus atributos, assim como todo crime possui muitos atributos, assim como todo jogo de basquete universitário possui seus muitos atributos.

Esses três exemplos são candidatos óbvios para estrutura, principalmente por sua ubiquidade. Pessoas têm destrinchado estatísticas esportivas por anos. Pessoas têm analisado crimes por anos.

Mas não precisamos parar nesses exemplos óbvios. Se você tirar um tempo para examinar que tipo de informações jornalistas de jornais coleta, uma grande quantidade de

informações estruturadas pulará em você (HOLOVATY, 2006)⁴⁴.

O pensamento de Adrian Holovaty vai de encontro com a proposta de produção do *data-driven journalism* descrita anteriormente e, principalmente, ao papel cada vez maior que bases de dados e informações abertas ao público – científicas ou governamentais – passaram a ocupar na produção jornalística na década de 2000. Apesar de não ser nova, a ideia de "open data" – ou seja, de que determinados dados devem ser disponíveis ao público de graça, de maneira acessível e com a possibilidade de serem reutilizados e redistribuídos – tem ganhado força e se consolidado com o crescimento da Internet e o surgimento de inúmeras organizações e iniciativas a favor da causa – como a *The Open Knowledge Foundation*⁴⁵, o *Open Data Manual*⁴⁶, o *Data.gov*⁴⁷, o *Data.gov.uk*⁴⁸, o *Portal da Transparéncia do Governo Federal*⁴⁹ e as *Páginas da Transparéncia Pública*⁵⁰. Este movimento também pode ser notado dentro das próprias organizações jornalísticas devido à multiplicação de seções, páginas e blogs mantidos por grandes jornais e que são voltados especificamente para as

⁴⁴ Disponível em: <http://www.holovaty.com/writing/fundamental-change/>. Acessado em: 12/10/2011. No original: "For example, say a newspaper has written a story about a local fire. Being able to read that story on a cell phone is fine and dandy. Hooray, technology! But what I really want to be able to do is explore the raw facts of that story, one by one, with layers of attribution, and an infrastructure for comparing the details of the fire -- date, time, place, victims, fire station number, distance from fire department, names and years experience of firemen on the scene, time it took for firemen to arrive -- with the details of previous fires. And subsequent fires, whenever they happen.

That's what I mean by structured data: information with attributes that are consistent across a domain. Every fire has those attributes, just as every reported crime has many attributes, just as every college basketball game has many attributes.

Those three examples are obvious candidates for structure, mostly due to ubiquity. People have been slicing and dicing sports stats for years. People have been analyzing crime for years.

But it doesn't stop at those obvious examples. If you take some time to examine what sort of information newspaper journalists collect, the amount of structure will jump at you".

⁴⁵ Fundada em 2004, é uma organização sem fins lucrativos que promove o conhecimento aberto, ou seja, qualquer tipo de informação que possa ser livremente usada, reusada e redistribuída (<http://okfn.org>).

⁴⁶ Desenvolvido pela Open Knowledge Foundation, é um manual disponibilizado online com discussões a respeito de aspectos legais, sociais e técnicos sobre dados abertos (<http://opendatamanual.org/#>).

⁴⁷ Site do governo norte-americano lançado em maio de 2009 com o propósito de aumentar o acesso público de alto nível a dados gerados pelo Poder Executivo do Governo Federal que não possuam restrições privadas ou ofereçam risco a segurança nacional (<http://www.data.gov/>).

⁴⁸ Tendo como um dos responsáveis Tim Berners-Lee, o site britânico possui um caráter semelhante ao *Data.gov*: seu objetivo é: a disponibilização pública de dados do governo britânico (<http://data.gov.uk>).

⁴⁹ Lançado em 2004, é o portal criado pelo governo federal brasileiro em conformidade com a Lei Complementar 131, de 2009, e que apresenta ao cidadão brasileiro informações sobre a aplicação de recursos públicos federais a partir de dados oriundos de diversos órgãos do Governo Federal relativos a Programas e Ações de Governo (<http://www.portaldatransparencia.gov.br>).

⁵⁰ Site do governo federal brasileiro que apresenta dados referentes às despesas realizadas por cada órgão e entidade da Administração Pública Federal (<http://www3.transparencia.gov.br/TransparenciaPublica/>).

discussões e produções envolvendo bases de dados, livres ou não: por exemplo, o jornal britânico *The Guardian* mantém uma página exclusiva para a sua produção baseada em dados, o *Data Store*⁵¹, e um blog para discutir as tendências do *data-driven journalism* e de iniciativas de dados livres, o *Datablog*⁵²; a equipe de desenvolvimento do *The New York Times* mantém um blog sobre desenvolvimento e que frequentemente aborda a questão de dados livres e a utilização de bases de dados na produção jornalística, o *Code*⁵³; o *Los Angeles Times* mantém uma página exclusiva para seus projetos com visualização de dados, o *Data Desk*⁵⁴; entre outros. Entretanto, para o infografista Alberto Cairo, a consolidação da tendência de transparência de dados é algo que vai além das iniciativas dos próprios governantes e dos meios jornalísticos. É algo que deve partir da conscientização de cidadãos.

"Isso sim é uma tendência que eu acho que vai continuar e vai melhorar se o cidadão em geral ficar ciente que isso é importante para ele. E se ele e os governantes tomarem consciência de que os dados que o governo tem não são propriedades do governo, mas sim dos cidadãos. Isso é uma mentalidade que, sobretudo nos países de origem latina, como o Brasil, precisa se firmar ainda. A mentalidade vigente é que se o governo disponibiliza esses dados, ele está fazendo um favor. O que não é correto já que quando você solicita isso, você está pedindo algo que pertence a você, algo a que você tem direito e algo que é obrigatório, ou deveria ser obrigatório, ao governo fornecer. E não só fornecer esses dados, mas fornecê-los em formatos flexíveis, que é algo que no Brasil precisa avançar um pouco já que os dados são fornecidos normalmente em formato pdf, que são difíceis de lidar. Então é uma tendência que vai continuar? Eu acredito que sim. Mas desde que não pare de se ter uma pressão por essa tendência. Uma pressão da sociedade, de movimentos como o de "open data", de hackers, etc" (CAIRO, 2011)⁵⁵.

⁵¹ <http://www.guardian.co.uk/data>

⁵² <http://www.guardian.co.uk/news/datablog>

⁵³ <http://open.blogs.nytimes.com/>

⁵⁴ <http://projects.latimes.com/index/>

⁵⁵ Alberto Cairo, diretor de infografia e multimídia da Revista Época, em entrevista realizada pelo autor em 26/11/2011.

Conforme Mirko Lorenz comentou em Amsterdã, essa tendência, aliada ao desenvolvimento de ferramentas de acesso a dados mais amigáveis, é uma das principais mudanças a afetar o jornalismo nos últimos anos e um dos pilares do *data driven journalism*.

Há dez, até cinco anos atrás, o uso de dados como base para reportagens era difícil e custoso, requerendo habilidades em TI muito além das comuns nos veículos de mídia. Bases de dados eram usadas principalmente por jornalistas investigativos. Editores e repórteres frequentemente se apoiavam sobre informações providas por fontes externas.

Hoje, há uma notável mudança. Coleções de dados estão se tornando disponíveis online, frequentemente de graça. Há uma pilha de ferramentas para se cavar "grandes dados". Ferramentas de código livre permitem a navegação e a análise de grandes quantidades de dados rapidamente. Há aplicações online que permitem compartilhar e visualizar dados (LORENZ, 2010: 06)⁵⁶.

Em seu artigo *Data as journalism, journalism as data*, de 2007, Rich Gordon, já apontava a tendência e a necessidade do aumento da utilização de dados na produção do jornalismo digital. Dentre as razões apontadas pelo autor do por que dados devem ser a estrutura guia do jornalismo online, destacam-se: dados são "conteúdo verde", ou seja, não perdem seu valor para o leitor após 24 horas; dados podem ser pessoais, dizendo respeito a problemas e situações locais; dados podem ser melhor entregues ao leitor em uma mídia sem restrições de espaço, como a Internet; dados podem tirar vantagem da maneira que as pessoas utilizam a Internet, já que são estruturas de comportamento ativo, ou seja, respondem às atividades dos leitores; e dados, após reunidos, podem ser compilados no impresso (GORDON, 2007).

⁵⁶ No original: "Ten, even five years ago, the use of data as a basis for reporting was difficult and costly, requiring IT skills far beyond what is common in the media. Databases were used mainly by investigative journalists. Editors and reporters usually relied on information provided by outside sources

Today there is a notable change. Collections of data are becoming available online, often for free. There is a whole stack of tools for digging into 'big data'. Open source tools allow navigation and analysis of large amounts of data rather quickly. There are online applications that allow us to share and visualize data" .

Ao tentar lançar luz sobre o *data-driven journalism*, procurando esmiuçar sua definição, características e técnicas, alguns dos participantes do seminário *Data-driven journalism: What is there to learn?* também falaram sobre a questão de dados livres e do que eles representam para a prática jornalística. Para Jonathan Gray, coordenador comunitário da *Open Knowledge Foundation*, a tendência da disponibilização de dados públicos nos levará a um ecossistema de informações livres e interligadas, caracterizadas por sua fácil reutilização, um grande número de contribuintes, uma distribuição descentralizada e por um caminho de inovação e novas formas de produzir e apresentar histórias. Por isso, ainda que esteja nos seus primeiros anos de vida, esse ecossistema tem o potencial para transformar radicalmente o modo como o conhecimento é disseminado em nossa sociedade. Gray acredita que, para os jornalistas, dados livres significam novas possibilidades de fazer e propagar notícias a partir do interesse público, de uma melhor contextualização das histórias e da publicação bases de dados com elas. Nesse contexto, ele reconhece que os jornalistas ainda têm um longo caminho a percorrer e que dados públicos ainda têm como principal desafio superar a incerteza legal de sua reutilização e redistribuição, já que muitas das iniciativas desse tipo ainda não especificam legalmente o que pode ou não ser feito com as informações disponibilizadas – tudo ainda é muito pressuposto (LORENZ 2010: 19-22).

O aumento da disponibilidade de dados livres e estruturados trouxe consigo crescentes preocupações de como melhor apresentar essas informações para o público a partir de estruturas gráficas, e como fazer isso sem que seja necessário ao jornalista um vasto conhecimento em linguagens de programação e outras tecnicidades. Ao mesmo tempo, conforme afirma Walter Wattenberg, pesquisador de visualização de dados da IBM, o desejo de se criar modelos de visualização de dados estimula o movimento para que essas informações sejam mais disponibilizadas de forma livre.

"Uma das coisas a respeito da visualização (de dados) é que ela definitivamente promove a abertura (de dados), pelo simples fato de existir, já que a partir do momento em que você começa a tentar trabalhar com visualizações, você precisa de dados. E se você tiver buracos nos seus dados, isso se tornará muito aparente. Dessa forma, a visualização possui dois papéis: um é bem explícito, que é o de nos deixar ver tendências; e o outro é esse papel meio passivo-agressivo de que, se você precisa

produzir modelos de visualização, você se encontrará inclinado a lutar pela causa dos dados livres" (WATTENBERG, 2009-2010)⁵⁷.

Dessa forma, é possível estabelecer que a relação entre o movimento de disponibilização de dados livres e as tentativas de visualizar esses dados pode ser estabelecida como um dos principais pilares do *data-driven journalism*.

4.3 - Ver para crer: a visualização de bases de dados no data-driven journalism

"Basicamente, metade de nosso cérebro é hardware para visão. Visão é um dos maiores benefícios que nós temos em termos de informações sensoriais para o mundo externo. Então o que a visualização (de dados) está fazendo é tirar vantagem do fato de sermos programados para entender o mundo ao nosso redor em termos daquilo que vemos" (VIEGAS, 2009-2010)⁵⁸.

"Parece que o mundo está sofrendo de excesso de informação ou abundância de dados. E a boa notícia é que pode haver uma solução fácil para isso, e é usarmos mais os nossos olhos. E assim visualizando informação, para que possamos ver os padrões e as conexões que importam, e então projetar a informação para que faça mais sentido, ou conte uma estória, ou nos permita focar apenas na informação que for importante. Se

⁵⁷ Martin Wattenberg, pesquisador de visualização de dados, em entrevista na vídeo-reportagem *Journalism at the age of data*, realizado entre 2009-2010. Disponível em: <http://datajournalism.stanford.edu/>. No original: "One of the things of about visualization is that it definitely can promote openness, just by the fact it existing, because the moment you try to make a visualization you realize you need data. And if you have holes in your data, it can become apparent. And so, visualization really concerns two rows: one is the explicitly row, of letting us see stuff coming; the other is this sort of passive-aggressive row that if you need to do a visualization you find yourself pushed to the goal of making data open"

⁵⁸ Fernanda Viegas, pesquisadora em visualização de dados, em entrevista na vídeore-portagem *Journalism at the age of data*, realizado entre 2009-2010. Disponível em: <http://datajournalism.stanford.edu/>. No original: "Basically, half of our brain is hardware for vision. Vision is the biggest benefit that we have in terms of sensory information to the outside world. So what visualization is doing is that is just taking advantage that we are so programmed to understand the world around us in terms of what we see"

não funcionar, a visualização de informação pode simplesmente parecer muito legal" (MCCANDLESS, 2010)⁵⁹.

Apesar da exploração geográfica, da matemática e da História terem estimulado a criação dos primeiros mapas, gráficos e linhas de tempo por volta de 1600, a forma atual mais popular de visualização de dados remonta há quase dois séculos, quando o engenheiro e político econômico escocês William Playfair (1759-1823) criou aquele que ficou conhecido como o primeiro gráfico com linhas e barras, em seu livro *Commercial and Political Atlas*, de 1786. Em 1801, Playfair repetiu seu pioneirismo, mas dessa vez publicando aquele que considerado como o primeiro gráfico de torta, no livro *Statistical Breviary*. Desde então, nomes como William Smith (1769-1839), Francis Galton (1822-1911), Florence Nightingale (1820-1910), Charles Minard (1781-1870), entre outros, continuaram a agregar evoluções aos modelos de Playfair e a criar novas formas de se visualizar informações. Em 1967, o cartógrafo e teórico francês Jacques Bertin (1918-2010) publicou o livro que é tido como a fundação teórica do modelo moderno de visualização de informações, o *Semiologie Graphique*. Nele, Jacques Bertin determina alguns padrões que até hoje são úteis em diversos trabalhos, como a definição de seis variáveis visuais que são diretamente aplicáveis a qualquer visualização gráfica (ALLEN, 2010).

Nas últimas duas décadas, a tecnologia e a Internet trouxeram consigo uma revolução para o mundo da visualização de dados ao torná-lo mais acessível e apresentar uma série de novas ferramentas e possibilidades, incorporando elementos como animação, interação, renderização de gráficos, alimentação de dados em tempo real, entre outros. Conforme Alberto Cairo disse em entrevista para este trabalho, nos últimos 15 anos a infografia deixou de ser algo estático para se tornar algo dinâmico e, principalmente, deixou de ser algo fechado em si, para se tornar algo mais atento às áreas complementares ao seu trabalho, como a cartografia, a computação etc. Para Shawn Allen, diretor de design de um estúdio especializado em visualização de dados e mapeamento em São Francisco, a Internet trouxe a possibilidade da visualização de

⁵⁹ David McCandless, jornalista, como parte da apresentação *The beauty of data visualization*, realizada em julho de 2010, TEDxGlobal 2010. Disponível em: http://www.ted.com/talks/lang/eng/david_mccandless_the_beauty_of_data_visualization.html No original: "It feels like we're all suffering from information overload or data glut. And the good news is there might be an easy solution to that, and that's using our eyes more. So, visualizing information, so that we can see the patterns and connections that matter and then designing that information so it makes more sense, or it tells a story, or allows us to focus only on the information that's important. Falling that, visualized information can just look really cool".

dados no jornalismo passar do status de um gráfico de barras em um jornal para "uma interface que permite contar a história". Ou seja, ela está deixando o papel de auxiliar na notícia para ocupar o papel de protagonista nela.

De repente, as pessoas parecem estar interessadas em dados; e esse interesse tem despertado uma necessidade de ferramentas visuais que os ajudem entendê-los. A visualização, em resposta a essa necessidade, tem se tornado cada vez dinâmica. Não é mais prático criar a maioria das tabelas e gráficos à mão. Ao invés disso, nós definimos novos padrões para valores dinâmicos em escala; novas interfaces para interativamente manipular dimensões em tabelas, como o tempo; e nós temos desenvolvido novas ferramentas para lidar com dados (ALLEN, 2010)⁶⁰.

Este fenômeno, vivido pelo jornalismo digital contemporâneo e descrito por Shawn Allen, foi mais bem analisado na vídeo-reportagem *Journalism in the Age of Data*, produzida entre 2009 e 2010 por Geoff McGhee, jornalista online especializado em multimídia e informação gráfica, com o auxílio do *John S Knight Journalism Fellowship*⁶¹. Dividido em sete capítulos, o vídeo de 60 minutos explora, a partir de diversas entrevistas com jornalistas, infografistas e pesquisadores a tendência da popularização da visualização de dados para os campos da arte e da comunicação, abordando tópicos como: a forma que a reportagem baseada em dados está mudando a estrutura das redações; o modo que jornalistas estão adaptando as narrativas tradicionais de reportagem a formatos baseados em visualização de dados; a efervescência do momento para os profissionais de infografia; como dados pode ser uma mídia expressiva; a importância da análise de dados para o bom jornalismo, as novas tecnologias e ferramentas que tornam a produção de visualizações de dados digitais acessíveis mesmo para aqueles profissionais que não são especialistas em infografia,

⁶⁰ No original: "People are, seemingly all of a sudden, interested in data; and that interest has in turn sparked a need for visual tools that help them understand it. Visualization, in response to this need, has become increasingly dynamic. It's no longer practical to create most charts or graphs by hand. Instead, we've designed new patterns for dynamic value scales; new interfaces for interactively manipulating chart dimensions, such as time; and we've developed new tools for managing data".

⁶¹ É o programa de bolsas para jornalistas da Universidade de Stanford, nomeado em homenagem ao editor de jornais americano John S. Knight.

estatística ou programação; e como jornalistas podem começar a produzir suas próprias visualizações de dados.

A partir da perspectiva de diferentes profissionais entrevistados ao longo da reportagem, fica clara a diferença entre a visualização de dados que era realizada na produção jornalística antes e a realizada agora, bem como importância do papel que a visualização de dados está passando a ocupar nas práticas jornalísticas contemporâneas, em um contexto em que o movimento pela disponibilização de informações e dados livres na Internet por empresas e organizações governamentais se intensifica cada vez mais.

Antes, com as estruturas de visualização de dados estáticas, os textos das produções jornalísticas contemplavam a história, enquanto os gráficos e tabelas se limitavam a oferecer evidências ou detalhes relacionados às narrativa. Agora, muitas das visualizações de dados produzidas no jornalismo, especificamente o online, tentam combinar narrativas e gráficos interativos de modo que estes sejam produtos jornalísticos autossuficientes para os leitores, e não apenas complementares. Para isso, ajuda a multiplicação de ferramentas online e gratuitas que tornam fácil a produção de visualizações de bases de dados, sem que seja necessário dos jornalistas vastos conhecimentos sobre programação e design. Dentre essas ferramentas, podem ser mencionadas algumas produzidas pela *Google*, como o *Google Docs and Spreadsheets*⁶², o *Google Fusion*⁶³ e o *Google Code Playground*⁶⁴; o *Yahoo Pipes*⁶⁵; o *ManyEyes*⁶⁶; o *Protopis*⁶⁷; entre outros.

Com o objetivo de investigar, classificar e definir características comuns às visualizações de dados narrativas produzidas no jornalismo, os pesquisadores Edward Segel e Jeffrey Heer, da Universidade de Stanford, realizaram 58 estudos de caso de produções online de jornais, como o *The New York Times*, *The Guardian* e o *The Financial Times*, blogs e pesquisas. As produções foram analisadas a partir do uso ou

⁶² Ferramenta disponibilizada pela *Google* para a produção e compartilhamento online de documentos, apresentações e tabelas - <http://docs.google.com>

⁶³ Ferramenta online disponibilizada pela *Google* para visualização e publicação de dados como mapas, linhas do tempo e gráficos - <http://www.google.com.br/fusiontables/Home/>

⁶⁴ Ferramenta online disponibilizada pela *Google* que oferece API's -- interfaces de programação de aplicações -- e recursos técnicos para desenvolvedores - <http://code.google.com/apis/ajax/playground>

⁶⁵ Ferramenta online disponibilizada pelo *Yahoo* para a agregação, manipulação e combinação de dados disponíveis na Internet - <http://pipes.yahoo.com/pipes/>

⁶⁶ Ferramenta online desenvolvida pela *IBM* para a produção de visualizações gráficas de dados - <http://www-958.ibm.com/software/data/cognos/maneyes/>

⁶⁷ Conjunto de ferramentas gráficas disponibilizado online para a visualização de dados - <http://mbostock.github.com/protovis/>

não de 16 elementos de narrativa em visualização (divididos em três categorias) e de 17 elementos de narrativa estruturada (também divididos em três categorias). A partir das análises, os pesquisadores identificaram o destaque de três importantes padrões nas produções: o agrupamento de diferentes estruturas de ordenação; a consistência de designs de interação; e a subutilização de mensagens narrativas. Além disso, encontraram sete gêneros básicos de visualizações narrativas que são usados na produção jornalística, de maneira individual ou combinados, e cuja adequação varia de acordo com as características e o objetivo da produção: estilo de revista, gráfico de anotações, poster particionado, gráfico de fluxo, tira, slide show e filme/vídeo/animação.

Mesmo que esses gêneros possam ser usados para contar uma história, nós notamos que diferentes gêneros funcionam melhor com diferentes tipos de histórias. Escolher apropriadamente o gênero depende de uma variedade de fatores, incluindo a complexidade dos dados, a complexidade da história, a audiência alvo e a mídia pretendida. Há claros casos em que um gênero é mais apropriado para um objetivo particular (...) Em geral, *a priori*, não há "resposta certa", mas diversas possibilidades candidatas, cada uma com suas vantagens e desvantagens (HEER e SEGEL, 2010: 8)⁶⁸.

Na produção jornalística, esses gêneros podem ser complementados com mensagens de texto, ou áudio, que oferecem observações ou explicações ao leitor sobre as imagens, e com recursos de interatividade, que permitem a manipulação dos dados e das imagens. Esta interatividade pode ser implementada em maior ou menor grau e com diferentes tipos de elementos, sendo os mais comuns botões de navegação, elementos que se iluminam quando o cursor passeia sobre eles, filtros de busca, barras de rolagem, capacidade de zoom in e zoom out e expansão com o tempo. Para Edward Segel e Jeffrey Heer, a partir de seu estudo, a narrativa dos produtos de visualização de dados se enquadra entre dois modelos chamados por eles de abordagens dirigidas ao autor

⁶⁸ No original: "Though each of these genres can be used to tell a story, we note that different genres work better for different story types. Choosing the appropriate genre depends on a variety of factors, including the complexity of data, the complexity of the story, the intended audience, and the intended medium. There are clear cases in which a genre is more appropriate for a particular purpose (...) In general, there will be no "right answer" *a priori*, but several possible candidates, each with advantages and disadvantages".

("author-driven") e abordagens dirigidas ao leitor ("reader-driven"). O primeiro se caracteriza por possuir um caminho linear na visualização de dados, contar com grande auxílio de mensagens e não possuir elementos de interatividade. Já o segundo não estabelece nenhuma ordem predefinida para a exploração na visualização, praticamente não faz uso de mensagens e ofereça grande liberdade de interação ao leitor. Diferentes produtos de visualização de dados possuem diferentes abordagens, mais ou menos dirigidas ao autor e ao leitor. Em seus estudos, os dois pesquisadores identificaram três modelos híbridos mais comuns de narrativa em visualizações de dados:

- **A estrutura de taça de martíni ("Martini Glass Structure")** - Esta estrutura começa com uma abordagem mais próxima da dirigida ao autor, com muita utilização de mensagens, mas depois se torna mais dirigida ao leitor, com a redução de mensagens e o aumento do nível de interatividade com o usuário para que este possa explorar os dados;
- **Slide–show interativo ("Interative Slide show")** - Se assemelha a uma apresentação de slides, mas incorpora recursos de narração interativa, permitindo ao leitor explorar determinados pontos da apresentação antes que avance na narrativa. Ao contrário da estrutura de taça de martíni, que se aproxima mais do modelo dirigido ao autor, o slide–show interativo realiza um balanço mais equilibrado entre as abordagens de autor e leitor;
- **História expansível ("Drible–Down Store")** - Apresenta uma estrutura de visualização baseada em um tema geral e permite ao leitor escolher entre desmembrações deste tema, revelando detalhes e particularidades que formam o todo. Ao oferecer uma liberdade maior de navegação, com opções, ao invés de uma narrativa linear, esta estrutura se aproxima mais do modelo dirigido ao leitor.

Acontece que, com as recentes possibilidades oferecidas pela Internet, pelas iniciativas de dados livres e diversas ferramentas de visualização de dados, surgem novos desafios e necessidades particulares às características do campo da visualização de informações, como as descritas por Nenhum Gershon e Ward Page.

Dados crus e informação são frequentemente complexos, em grande volume, dependentes do tempo, de diversos tipos e de diversas fontes, e nem sempre são confiáveis. Grandes quantidades de informação criam um problema de escala (como a representação de uma grande quantidade de informações simultaneamente em uma mesma tela). Métodos para lidar com questões de escalas incluem a organização, condensação, segmentação e sumarização de informações.

(...)

Por outro lado, usuários possuem diferentes níveis de habilidades pessoais, educação, e gostos (diferentemente da visualização científica, que é voltada para cientistas treinados).

(...)

Como sistemas de visualização podem ser moldados para acomodar a percepção e o processamento humano de informação? Os desenvolvedores de softwares de visualização e os designers de apresentação precisam entender como humanos interagem tanto visualmente como não visualmente e como percebem informações, assim como a mente humana trabalha quando busca por informações conhecidas ou desconhecidas e na solução de problemas. Mesmo que uma interação humana-computador seja central em visualizações, nem sempre é adequada por si só para fazer usuários entenderem o que eles estão olhando. Desenvolvedores de softwares e designers de apresentação também precisam programar o que nós sabemos sobre como humanos entendem e interagem com informações e nosso próprio sistema de percepção. Eles também precisam aprender como criar interfaces de usuário flexíveis, ferramentas de navegação e métodos de pesquisa apropriados para cada tipo de usuário, aplicação e tarefa (GERSHON, Nenhum e PAGE, Ward, 2001: 33)⁶⁹

⁶⁹ No original: "Raw data and information are often complex, high-volume, time-dependent, of diverse types from diverse sources, and not always reliable. Massive amounts of information create a problem of scaling (such as for representing a massive amount of information simultaneously on the same screen). Methods for dealing with scaling issues include information organization, condensation, segmentation and summarization.(...)

On the other hand, users come with all types and levels of personal skills, education, and tastes (unlike in scientific visualization, which is intended for highly trained scientists).(...)

Esses desafios encontram eco nas afirmações de Alberto Cairo feitas durante a entrevista para este trabalho.

"Os desafios batem um pouco com os passos do *data-driven journalism*. O primeiro desafio continua sendo conseguir os dados. Porque os dados nem sempre estão disponíveis, se estão disponíveis pode ser que não estejam no formato certo, às vezes são não contrastáveis pois são de só uma fonte e você tem que confiar nessa fonte. Então há o desafio de obtenção e tratamento de dados.

Agora, na hora de apresentar os dados, você tem vários desafios diferentes dependendo do tipo de gráfico que você esteja criando. O desafio principal é transmitir de forma gráfica uma visão das proporções, das relações, das causalidades, o que não é sempre fácil de obter e nem sempre pode ser conseguido de uma forma intuitiva. Muitas vezes, você acabar construindo um gráfico da forma errada. Por exemplo, o gráfico de bolhas: ele às vezes é usado em substituição ao gráfico de barras, o que é errado porque o gráfico de bolhas não deixa perceber as proporções de duas magnitudes do mesmo jeito que um gráfico de barras o faz, de forma precisa. Então há o desafio na visualização que o jornalista, ou o design que esteja fazendo o infográfico, se pergunte para que este gráfico tenha que servir, qual é o objetivo do gráfico.

O gráfico no final não é uma ilustração, não é um desenho. É uma ferramenta cognitiva que o leitor vai usar para interpretar e

How can visualization systems be tailored to accommodate human perception and information processing? Visualization software developers and presentation designers need to understand how humans interact both visually and nonvisually with and perceive information, as well as how the human mind works when searching for known and unknown information and solving problems. Even though effective human-computer interaction is central in visualization, it is not always adequate by itself for making users understand what they're looking at. Visualization software developers and presentation designers also need to implement what we know about how humans understand and interact with information and our built-in perceptual systems. They also need to learn how to create flexible user interfaces, navigation tools, and search methods appropriate for each type of user, application, and task".

analisar os dados. Então, como qualquer outra ferramenta, a forma que essa ferramenta adota deve ser adaptada às funções que ela tem. A escolha das formas gráficas não depende do gosto estético do designer, mas sim das funções que esse gráfico tem que ter, das respostas às perguntas que esse gráfico tenta oferecer" (CAIRO, 2011)⁷⁰.

⁷⁰ Alberto Cairo, diretor de infografia e multimídia da Revista Época, em entrevista realizada pelo autor em 26/11/2011.

5 - Estudo de caso: a cobertura feita pelo *The Guardian* sobre os *war logs* da Guerra do Afeganistão disponibilizados pela *Wikileaks*

5.1 - O que foi e o por que de sua escolha para este trabalho?

Em julho de 2010, o *Wikileaks*⁷¹, uma organização internacional sem fins lucrativos, levou a público cerca de 90 mil relatórios de inteligência e documentos do exército americano, classificados como sigilosos, sobre a ocupação dos Estados Unidos no Afeganistão durante a Guerra do Afeganistão e produzidos entre janeiro de 2004 e dezembro de 2009. Os documentos ficaram conhecidos como "*war logs* do Afeganistão".

A publicação desses documentos foi feita pelo *Wikileaks* em seu site oficial, mas também de forma simultânea e em parceria com três grandes empresas de mídia: o jornal britânico *The Guardian*, o jornal americano *The New York Times*, e a revista semanal alemã *Der Spiegel*. Os três veículos tiveram acesso aos documentos pelo *Wikileaks* semanas antes de sua publicação, para que pudessem confirmar sua autenticidade e certificar sua relevância. Em troca, os veículos tiveram total liberdade por parte do *Wikileaks* para trabalhar com os documentos e os publicar no formato e da maneira que bem quisessem, desde que os publicassem simultaneamente entre eles e o próprio *Wikileaks*.

Para além do significado e do impacto jornalístico e geopolítico que tiveram as revelações feitas pelo *Wikileaks* sobre a Guerra do Afeganistão, e para além das diferenças nas abordagens da temática e na apresentação dos documentos e arquivos pelos três veículos que puderam publicá-los em primeira mão, este trabalho se focará na análise da cobertura dos war logs feita pelo jornal britânico *The Guardian*, a partir dos conceitos e características sobre *data-driven journalism* e visualização de dados apresentados nos capítulos anteriores. A escolha pelo caso do vazamento de informações sigilosas sobre a Guerra do Afeganistão pelo *Wikileaks* como estudo de caso para este trabalho se dá, em particular, por sua relevância na cobertura jornalística

⁷¹ Publicamente notado em 2007, apesar de seu site ter ido ao ar em outubro de 2006 (<http://wikileaks.org/>), o *Wikileaks* se auto-denomina uma organização sem fins lucrativos cujo objetivo é "trazer importantes notícias e informações ao público", para que este possa ver "evidências da verdade". Nos últimos anos, o *Wikileaks* foi o responsável de uma série importantes documentos secretos do Exército norte-americano sobre a Guerra do Afeganistão, a Guerra do Iraque, além de documentos diplomáticos confidenciais de embaixadas norte-americanas ao redor do mundo contendo revelações e comentários sobre outros países.

no noticiário contemporâneo mundial, por ter ocorrido nos últimos cinco anos e pelas possibilidades ilustrativas que ele oferece aos temas do *data-driven journalism* e da visualização de dados. A escolha pela cobertura específica do jornal britânico *The Guardian* sobre o caso se dá pela importância que o veículo tem atribuído nos últimos anos às possibilidades oferecidas pelo *data-driven journalism* em sua produção e, principalmente, pelos usos que fez delas na cobertura dos war logs do Afeganistão.

5.2 - Como foi a cobertura do *Guardian*?

A cobertura completa do jornal britânico *Guardian* sobre os war logs do Afeganistão foi publicada no dia 25 de julho, por meio de uma página especial⁷² no site do jornal. O material produzido pelo *Guardian* em um momento inicial contou com: um infográfico interativo mostrando o aumento dos 16.000 registros do exército americano de ataques de *improvised explosive device* (IED) – bombas de fabricação caseira para estradas que são imprevisíveis e impossíveis de lutar – de janeiro de 2004 a janeiro de 2010, contendo detalhes como o tipo de IED, localização dos ataques e a quantidade de casualidades – separadas por tipos; um guia interativo para mais de 2.000 eventos registrados nos war logs e que incluem detalhes de situações como fogo amigo, casos onde soldados do exército americano atiraram ou mataram civis, e situações em que as forças afegãs dispararam umas contra as outras; a disponibilidade aos leitores dos dados crus em planilhas oferecidos pelo *Wikileaks*; além de pelo menos 11 textos e análises sobre o que as revelações dos documentos significam, sobre a Guerra de Afeganistão e seus envolvidos, e sobre como foi feita a própria cobertura do caso pelo *Guardian*.

Conforme descrito em um texto editorial publicado dois dias após a cobertura dos war logs⁷³ ter ido ao ar no site do jornal e ter sido publicada no impresso, toda a preparação do material jornalístico do *Guardian* sobre os documentos liberados pelo *Wikileaks* demorou semanas, teve como objetivo confirmar a veracidade das informações, destrinchar seu significado, mas também preparar os dados para que pudessem ser trabalhados de maneira mais fácil pelo time de repórteres investigativos do jornal – constituído por David Leigh, Nick Davies, Declan Walsh, Simon Tisdall e Richard Norton-Taylor – e, principalmente, tornar o resultado facilmente explorável e

⁷² <http://www.guardian.co.uk/world/the-war-logs>

⁷³ Disponível em: <http://www.guardian.co.uk/news/datablog/2010/jul/27/wikileaks-afghanistan-data-datajournalism>. Acessado em: 15/11/2011.

entendível para o leitor, de uma maneira que esse pudesse ter a ideia exata do significado do que antes eram apenas informações em uma planilha. De acordo com o *Guardian*, os dados do *Wikileaks* foram entregues ao jornal na forma de um arquivo de planilha contendo mais de 92.201 linhas de informação – algumas com nada nelas, resultado de má formatação – o que, somente pelo tamanho, representou um grande desafio técnico para o jornal.

Qualquer coisa acima de 60.000 linhas traz problemas dramáticos para o Excel – salvar toma um tempo dolorosamente longo (dica número um – desligar os salvamento automático nas configurações...). O que não ajuda os repórteres a tentar trilhar um caminho sobre os dados na busca por histórias e é grande demais para se gerar relatórios significativos em cima dos dados. (ROGERS, 2010)⁷⁴.

Com isso, foi necessário que o editor de sistemas do *Guardian*, Harold Frayman, junto com John Houston, construísse uma base de dados nova com os dados da tabela entregue pelo *Wikileaks* para que os repórteres pudessem realizar buscas por eventos chaves e significativos. Ou seja, foi necessário transformar os dados crus em algo que pudesse ser trabalhado a partir de uma boa estrutura categorizada. Mas como verificar a veracidade das informações? Para isso o *Guardian* cruzou os dados do *Wikileaks* com relatórios de casualidades da Organização do Tratado do Atlântico Norte (Otan). Uma vez tornado os dados trabalháveis, o desafio foi filtrar as boas histórias e informações relevantes, e, por meio das possibilidades gráficas oferecidas por recursos de visualização de dados – como mapas, infográficos, gráficos, tabelas, símbolos e cores –, apresentar as informações aos leitores, no impresso e na Internet, de um modo que esses pudessem ter a real dimensão do significado da situação vivida por soldados e afegãos durante a Guerra do Afeganistão. Online, o desafio foi não só esse, mas também o de tornar esses dados navegáveis para os leitores, para que eles pudessem tirar suas próprias conclusões a partir de suas próprias experiências com as informações.

A cobertura do *Guardian* sobre os *war logs* da Guerra do Afeganistão é emblemática como estudo de caso para este trabalho porque embarca muitas das

⁷⁴ No original: "Anything over 60,000 rows or so brings excel down in dramatic fashion – saving takes a painfully long period of time (tip number one – turn automatic saving off in preferences...). It doesn't help reporters trying to trawl through the data for stories and it's too big to run meaningful reports on".

características descritas anteriormente por jornalistas e especialistas para as produções de *data-driven journalism*. O *Guardian* teve como base para a sua produção um conjunto de dados, disponibilizados em uma estrutura de base de dados – uma planilha, ainda que precariamente estruturada – pela *Wikileaks*, uma organização sem fins lucrativos cujo objetivo é a transparência de informações de governos e empresas para o público, semelhante ao movimento de "open data" pelo qual o *data-driven journalism* tem se fortalecido. É curioso observar ainda que toda a descrição do processo produtivo do *Guardian* oferecida pelo próprio jornal sobre os *war logs* da Guerra do Afeganistão condiz com as etapas de produção do *data-driven journalism* descritas anteriormente por Mirko Lorenz e Paul Bradshaw e que podem ser resumidas em: encontrar os dados, interrogar os dados, produzir visualizações desses dados, e deles extrair histórias. Fica assim claro como a cobertura dos *war logs* pelo *Guardian* pode ser considerado um produto legítimo de *data-driven journalism* – algo reconhecido por Simon Rogers, editor de notícias e de duas seções do *Guardian* dedicadas ao jornalismo de dados: o blog *Datablog* e a página *Datastore*.

Bem, nós sempre quisemos histórias a partir de dados: agora nós conseguimos. Em grande quantidade. Com sinos nelas. Os *war logs* sobre o Afeganistão do *Wikileaks* são uma fantástica vitória para o jornalismo investigativo de dados, não apenas aqui no *Guardian*, mas também no *New York Times* e no *Der Spiegel*.

Também é o jornalismo de dados em ação. O que nós queríamos fazer era permitir que nosso time de repórteres especialistas tirassem ótimas histórias humanas da informação - e nós queríamos analisá-las para chegar ao contexto maior, para mostrar como a guerra realmente está acontecendo (ROGERS, 2010)⁷⁵.

⁷⁵ Retirado do artigo *Wikileaks' Afghanistan war logs: how our datajournalism operation worked*. Disponível em: <http://www.guardian.co.uk/news/datablog/2010/jul/27/wikileaks-afghanistan-data-datajournalism>. Acessado em: 10/11/2011. No original: "Well, we always wanted stories from data: now we've got it. In spades. With bells on. The Wikileaks' Afghanistan war logs are a fantastic victory for investigative data based journalism, not only here at the Guardian but at the New York Times and Der Spiegel too.

Com o objetivo de reforçar o trabalho do *Guardian* como sendo um produto de *data-driven journalism* e exemplificar as possibilidades oferecidas pela etapa de visualização de dados no processo produtivo, continuarei abaixo a análise da cobertura dos *war logs* do jornal com foco em seus infográficos.

5.3 - Os *war logs* do Afeganistão: a seleção do *Guardian*



Um dos infográficos construídos pelo *Guardian* foi um mapa com uma seleção dos 300 registros mais relevantes, do ponto de vista do jornal e dentro de um universo de 92.201 registros vazados pelo *Wikileaks*, mostrando a localização de cada um deles e suas respectivas classificações em categorias: acidente, fogo amigo das tropas afegãs, fogo amigos das forças de coalizão, casualidades civis, prisões e outros⁷⁶. Cada ocorrência é representada no mapa por um círculo de uma cor distinta, que varia de acordo com sua classificação, e, ao passar o cursor em cima de cada uma delas, é

It's also datajournalism in action. What we wanted to do was enable our team of specialist reporters to get great human stories from the information – and we wanted to analyse it to get the big picture, to show how the war really is going".

⁷⁶ Disponível em: <http://www.guardian.co.uk/world/datablog/interactive/2010/jul/25/afghanistan-war-logs-events>. Acessado em: 10/11/2011.

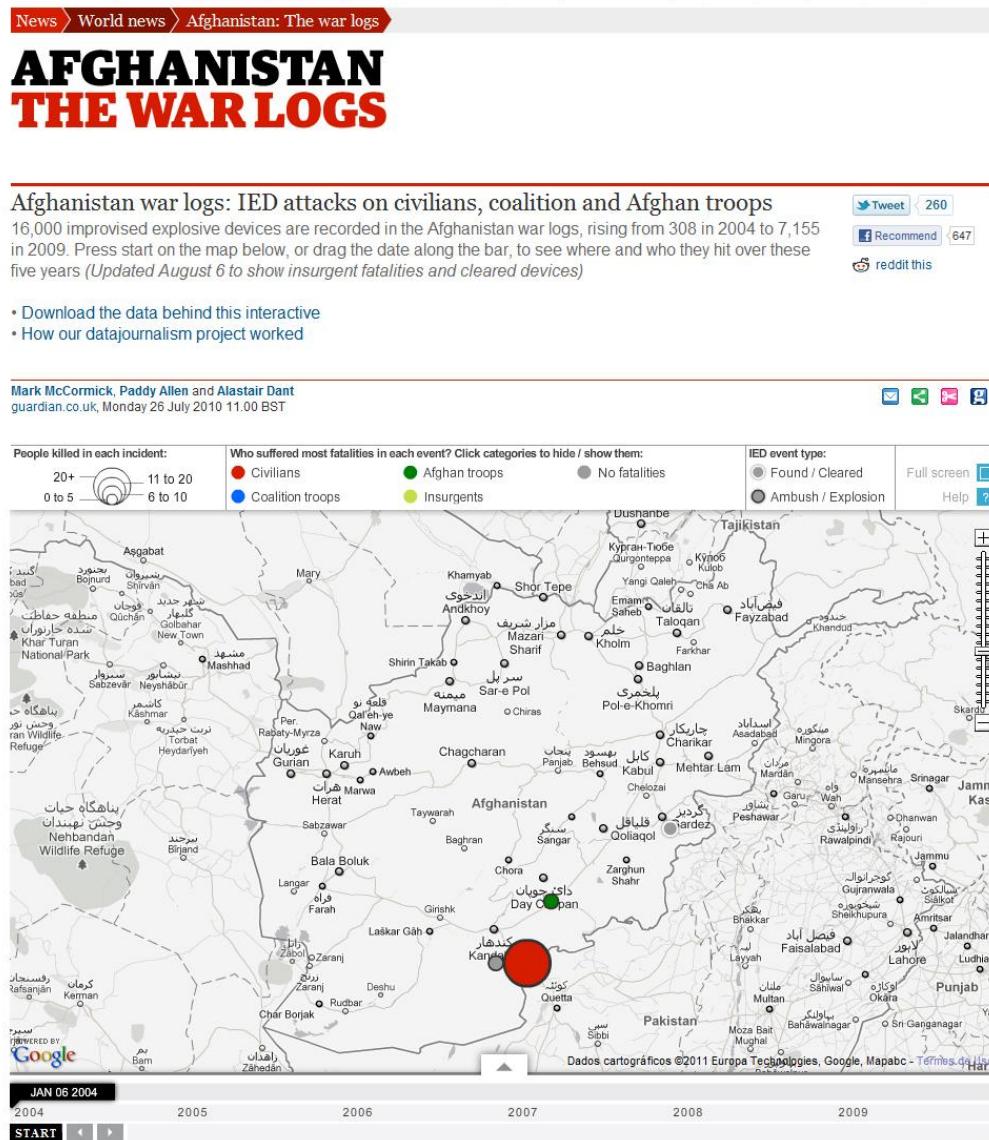
mostrado a data e a hora em que aconteceu. As possibilidades de interação com o mapa oferecidas pelo jornal são três: a capacidade de dar zoom-in ou zoom-out; a possibilidade de "acender" ou "apagar" determinada categoria de ocorrência do mapa e, dessa forma, visualizar todos os pontos coloridos ou fazer apenas com que o mapa só mostre, por exemplo, os eventos com casualidades civis; e verificar, ao passar o cursor por cima de cada um dos pontos coloridos, a data e o horário de sua ocorrência.

Utilizando os gêneros e parâmetros selecionados por Edward Segel e Jeffrey Heer em suas análises de produtos de narrativas visuais, podemos classificar este produto do *Guardian* como pertencendo ao gênero de gráfico de anotações/mapa e observar que ele faz uso alguns elementos de narrativa visual e de narrativa estruturada, conforme descrito abaixo:

- Elementos de narrativa visual utilizados:
 1. Estrutura visual: uma plataforma visual consistente;
 2. Elementos de destaque: zoom;
- Elementos de narrativa estruturada utilizados:
 1. Ordenação: acesso randômico;
 2. Interatividade: destaque pelo cursor; detalhes; filtragem de dados; seleção; instruções explícitas;
 3. Mensagens: legendas/ cabeçalhos; texto de introdução;

Dentro da descrição de leveis do jornalista Rich Gordon que um trabalho de jornalismo de dados pode alcançar, esta infografia do *Guardian*, quando analisada de forma isolada do especial sobre os *war logs* do jornal, vai até o nível da visualização de dados, o quarto nível de um total de cinco. O mapa, ainda que de modo óbvio, organiza as informações dos *war logs* de uma maneira que o usuário possa visualizá-las, ter algum grau de naveabilidade por elas e, a partir disso, ter uma compreensão maior de seus significados e da real dimensão da Guerra do Afeganistão, com um retrato do conflito mais fidedigno. No entanto, o mapa não constitui por si só uma experiência narrativa e nem apresenta, de modo isolado, um conteúdo jornalístico autosuficiente e completo: para a sua total compreensão e para que suas informações tomem uma proporção mais adequada é necessário ao leitor acompanhar o restante da cobertura do *Guardian* sobre os *war logs*.

5.4 - Os *war logs* do Afeganistão: os ataques IED em civis, forças da coalizão e tropas de Afegãos



A segunda peça infográfica produzida pelo *Guardian* em sua cobertura *dos war logs* da Guerra do Afeganistão foi um mapa interativo que mostra todos os 16.000 registros de incidentes com *improvised explosive devices* (IED) – bombas de fabricação caseira para estradas que são imprevisíveis e impossíveis de lutar – entre janeiro de 2004 e janeiro de 2010, com informações como a localização dos incidentes, os grupos mais afetados em cada incidente (civis, tropas da coalizão, tropas de afegãos, insurgentes), o número de pessoas mortas em cada incidente, e o tipo de incidente IED

(encontrado/limpo ou emboscada/explosão)⁷⁷. O mapa conta ainda em sua parte de baixo com uma barra de navegação cronológica. Nela, o leitor pode selecionar uma data específica para ver as informações sobre os incidentes daquele dia, ou movimentar a barra de navegação manualmente para acompanhar o crescimento ou diminuição dos incidentes no decorrer do tempo. É possível ainda estabelecer para a barra se movimentar de maneira automática do início ao fim do período registrado, apresentando uma animação dos incidentes no decorrer do tempo e com suas localizações no mapa.

De maneira semelhante ao mapa dos 300 *war logs* selecionados pelo *Guardian*, esta peça se classifica, dentro dos gêneros destacados por Edward Segel e Jeffrey Heer, como um gráfico de anotações/mapa. Entretanto, ao ter uma barra de navegação cronológica, ele incorpora mais elementos de interação e narrativa que o primeiro mapa analisado.

- Elementos de narrativa visual utilizados:
 1. Estrutura visual: plataforma visual consistente; barra de progressão/barra de tempo;
 2. Elementos de destaque: zoom; movimento;
 3. Guia de transição: transição animada;
- Elementos de narrativa estruturada:
 1. Ordenação: acesso randômico; caminho direto do usuário;
 2. Interatividade: destaque pelo cursor; detalhes; filtragem de dados; seleção; instruções explícitas; botões de navegação;
 3. Mensagens: legendas/ cabeçalhos; texto de introdução;

Apesar de esta peça ainda não constituir uma experiência de reportagem autosuficiente, ela agrupa elementos de narratividade em sua estrutura, o que a torna um produto de nível cinco na classificação de Rich Gordon. Ainda que o usuário dependa de outras informações fora do infográfico para ter uma melhor percepção sobre a Guerra do Afeganistão e as informações contidas nos *war logs*, no que diz respeito a informações sobre os incidentes com IED, este mapa pode ser considerado autosuficiente, constituindo uma experiência mais rica e completa ao usuário que o primeiro mapa analisado. Do ponto de vista narrativo, este mapa pode-se afirmar que

⁷⁷ Disponível em: <http://www.guardian.co.uk/world/datablog/interactive/2010/jul/26/ied-afghanistan-war-logs>. Acessado em: 10/11/2011.

este mapa utilizado algo semelhante ao modelo de Slide-show interativo ("Interactive Slideshow"), mencionado por Edward Segel e Jeffrey Heer, em que o mapa pode ser considerado uma apresentação de slides, em que cada slide seria uma determinada data, que oferece ao usuário determinados elementos de interação antes que avance com a apresentação.

6 - Considerações Finais

Atualmente, vivemos em uma realidade caracterizada pela explosão da oferta de informações e por um acelerado ritmo de vida – ao menos nos grandes centros urbanos. Parte desse ritmo é resultado do desenvolvimento tecnológico dos meios de comunicação, de transporte, das interfaces digitais, e das possibilidades proporcionadas pelo crescimento e pela popularização da Internet – um ambiente de fluxos de informações cada vez mais interativos, colaborativos, dinâmicos e velozes. Neste contexto, em meio à consolidação do jornalismo digital, as bases de dados se firmaram como peças chave na estruturação e administração de veículos produtores de notícias no ambiente online e, nos últimos anos, como os principais ingredientes de novas formas de produção de conteúdos condizentes com as características do paradigma do Jornalismo Digital de Base de Dados (JDBD).

Essas condições – aliadas a fatores como a multiplicação de iniciativas e organizações que lutam pelos valores do "open data" na Internet, e de ferramentas sofisticadas de visualização de dados, porém de fácil usabilidade – acabaram por dar origem ao *data-driven journalism*, tido como um dos mais promissores campos do jornalismo a emergir neste início de século XXI. Com este trabalho, a apreensão dos contextos e conceitos aqui descritos, e as análises realizadas, espero ter conseguido deixar claro para o leitor todos os fatores e condições que propiciaram a formação do *data-driven journalism*, suas características, particularidades, e um vislumbre das potencialidades e do por que da importância atribuída ao *data-driven journalism*. É importante compreender que o *data-driven journalism* não se propõe a revolucionar ou salvar o jornalismo, mas sim oferecer novas possibilidades de produção aos jornalistas e de produtos aos leitores. Antes de tudo, o *data-driven journalism* é uma ferramenta que ainda se ampara em habilidades tradicionais do jornalismo.

Conforme Alberto Cairo afirmou em entrevista para o autor, o *data-driven journalism* é um campo que ainda está em desenvolvimento, cujas possibilidades ainda estão aparecendo e cujos desafios ainda são primários: como a disponibilização de dados por governos e empresas em um formato adequado para que jornalistas possam trabalhar com eles; a falta nas redações de jornalistas que saibam das possibilidades que as bases de dados têm a oferecer e que possuam conhecimentos especializados em estatística e que saibam mexer com números; e a falta de formação técnica dos próprios designers, que, em muitos casos, pouco têm cultura jornalística ou de estatística.

Por ainda ser uma novidade, o *data-driven journalism* encontra uma disseminação tímida de suas práticas em muitas redações jornalísticas, e isso também se reflete na ausência de análises acadêmicas sobre o assunto, principalmente aqui no Brasil, onde sua prática é muito incipiente – ainda não há distinção entre produtos infográficos e produtos de *data-driven journalism* no país e a maioria deles ainda são gerados em torno de mapas, por exemplo. Dessa forma, este trabalho não tem a ambição de esgotar as possibilidades de análise e discussão sobre o tema, mas sim servir como uma referência, ou um ponto inicial para análises futuras, que podem se focar em produtos de *data-driven journalism* específicos, processos de produção dentro de veículos jornalísticos específicos, ou, ainda, análises na produção de grupos de mídia específicos, como os veículos jornalísticos brasileiros. E esses estudos somente poderão ser realizados à medida que o *data-driven journalism* se firmar como uma prática do jornalismo para além de uma simples tendência.

7 - Referência Bibliográfica:

ARMAÑANZAS, E., MESO, K., & NOCI, J. D. **El periodismo electrónico.** Información y servicios multimedia en la era del ciberespacio. Barcelona: Comunicación, 1996.

BALDESSAR, M. J. **A mudança anunciada:** o cotidiano dos jornalistas com computador na redação. Florianópolis: Insular, 2003.

____ **Jornalismo e tecnologia: pioneirismo e contradições na chegada da informatização nas redações catarinenses.** Revista PJ:Br (São Paulo), São Paulo, n. 5, 2005.

BARBOSA, S. **Jornalismo Digital em Base de Dados (JDBD)** - Um paradigma para produtos jornalísticos digitais dinâmicos. Tese apresentada no Curso de Doutorado em Comunicação e Cultura Contemporânea, da Faculdade de Comunicação da Universidade Federal da Bahia, 2007.

COLLE, R. **Explotar la información noticiosa.** Data mining aplicado a la documentación periodística. Madrid: Departamento de Biblioteconomía y Documentación, Universidad Complutense de Madrid, 2002.

COLOMBO, F. **Os arquivos imperfeitos.** São Paulo: Perspectiva, 1991.

FERRARI, P. **Jornalismo digital.** São Paulo: Contexto, 2003.

FIDALGO, A. **Do poliedro à esfera: os campos de classificação.** A resolução semântica no jornalismo online. In: *Anais do II SBPJor*. Salvador, 2004.

____ **Sintaxe e semântica das notícias on-line.** Para um jornalismo assente em base de dados. In: *FIDALGO, António; SERRA, Paulo (Orgs). Informação e Comunicação Online. Jornalismo Online. Volume 1*. Covilhã: Universidade da Beira Interior, 2003.

GARRISON, B. **Computer-Assisted Reporting.** London: LEA Publishers, 1998.

GERSHON, N., & PAGE, W. **What storytelling can do for information visualization.** *Communication of the ACM* , Vol. 44, Nº8, p. 31-37, 2001.

GUIMARÃES, C. C. (2003). **Fundamentos de bancos de dados.** Modelagem, projeto e linguagem SQL. Campinas : Unicamp, 2003.

GUNTER, B. **News and the net. New Jersey.** London: Lawrence Erlbaum Associates, Inc, 2003.

HOUSTON, B. **Computer-Assisted Reporting.** A practical guide. Boston/New York: Bedfort St. Martin's, 2004.

LIMA JÚNIOR, W. T. **Classificação das bases conceituais para validação de instâncias e atributos a serem utilizados na elaboração de sistemas digital de busca de fontes jornalísticas.** In: *Anais do IV SBJor (CD-ROM)*. Porto Alegre-RS/Brasil, 2006.

_____. Pesquisa Aplicada na descoberta de novas tecnologias de produção e distribuição de conteúdos jornalísticos. In: *XXVIII Congresso Brasileiro de Ciências da Comunicação*. Rio de Janeiro-RJ/Brasil, 2005.

_____. Jornalismo Inteligente (JI) na era do data mining. In: *Anais do II SBPJor*. Salvador-BA/Brasil, 2004.

LORENZ, M. **Data driven journalism: What is there to learn?** Amsterdã: European Journalism Centre (EJC), 2010.

MACHADO, E. **O Jornalismo Digital em Base de Dados.** Florianópolis: Calandra, 2006.

_____. **Banco de dados como formato no jornalismo digital.** *Anais III Sopcom, VI Lusocom, II Ibérico, UBI (CD-ROM)*, 2004.

____ **O banco de dados como espaço de composição de narrativas multimídias.** In: *Anais do II SBPJor*. Salvador, 2004.

MANOVICH, L. **The Language of new media.** Cambridge: MIT Press, 2001.

MEYER, P. ***The future of CAR: Declare Victory and Get Out! From "When Nerds and World Collide: Reflections on the Development of Computer Assisted Reporting.*** The Poynter Institute for Media Studies.

____ **The new precision journalism.** Bloomington: Indiana University Press, 1991.

MIELNICZUK, L. (2003). **Jornalismo na Web:** uma contribuição para o estudo do formato na escrita hipertextual. Tese apresentada ao Curso de Doutorado em Comunicação e Cultura Contemporânea, Faculdade de Comunicação da Universidade Federal da Bahia. Salvador, 2003.

PALACIOS, M. **Fazendo jornalismo em redes híbridas.** Notas para discussão da internet enquanto suporte mediático. Artigo produzido para discussão na Lista JnCultural. Minas Gerais, 2002.

PAUL, N. **Computer-Assisted Research.** A Guide to Tapping Online Information. Chicago: Bonus, 1999.

QUADROS, C. I.. **Base de dados: a memória extensiva do jornalismo.** Porto Alegre: In: Em Questão, V.11. N.2, Jul/dez., 2005.

ROSA, V. G. **O impacto das tecnologias nas redações:** como a informatização modificou a rotina profissional dos jornalistas. Trabalho de conclusão de curso de graduação em Comunicação Social, Universidade Federal de Juiz de Fora., 2005.

SCHWINGEL, C. **Jornalismo digital de quarta geração:** a emergência de sistemas automatizados para o processo de produção industrial no jornalismo digital. In: *Anais do XIV Encontro Anual da Compós*, 2005.

SEGEL, E., & HEER, J. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics*, v.16 n.6, p. 1139-48, 2010.

SILVA JR., J. A. **A relação das interfaces enquanto mediadoras de conteúdo do jornalismo contemporâneo**: agências de notícias como estudo de caso. *Trabalho apresentado no XI Encontro Anual da Compós*. Rio de Janeiro, 2002.

SMITH, A. **Goodbye Gutenberg**: The newspaper revolution of the 1980s. New York: Oxford University Press, 1980.

SQUIRRA, S. **Jorn@alismo Online**. São Paulo: Arte & Ciência, 1998.

Sites acessados:

ALLEN, S. *Introduction*. Disponível em:

<http://interactiondesign.sva.edu/classes/datavisualization/2010/07/08/introduction/>.

Acessado em: 16/09/2011

BARDOEL, J. &. *Network journalism: converging competences of old and new media professionals*. Disponível em: <http://home.pscw.uva.nl/deuze/pub19.htm>.

Acessado em: 16/09/2011

BRADSHAW, P. *How to be a data journalist*. Disponível em:

<http://www.guardian.co.uk/news/datablog/2010/oct/01/data-journalism-how-to-guide>.

Acessado em: 10/10/2011.

GORDON, R. *Data as journalism, journalism as data*. Disponível em

<http://getsmart.readership.org/2007/11/data-as-journalism-journalism-as-data.html>.

Acessado em: 10/10/2011.

HOLOVATY, A. *A fundamental way newspaper sites need to change*. Disponível em:

<http://www.holovaty.com/writing/fundamental-change/>. Acessado em: 25/10/2011.

____ *Dynamic News Stories*. Disponível em:
<http://www.xml.com/pub/a/2006/05/17/dynamic-news-stories.html>. Acesso em: 25/10/2011.

____ *Microformats could describe online news intelligently*. Disponível em:
<http://www.holovaty.com/writing/299/>. Acessado em: 25/10/2011.

MCADAMS, M. *The sad story of videotext*. Disponível em:
<http://www.well.com/user/mmcadams/videotext.html>. Acessado em: 06/11/2011

PRYOR, L. *The third wave of online journalism*. Online Journalism Review. Disponível em: <http://www.ojr.org/ojr/future/1019174689.php>. Acessado em 08/08/2011.

SCHEDDEN, D. *New Media Timeline (1969-1998)*. Disponível em:
<http://www.poynter.org/latest-news/business-news/transformation-tracker/28803/new-media-timeline-1969-2010/>. Acessado em: 07/08/2011

8 - Anexos:

8.1 - Lista de exemplos de produções de *data-driven journalism*

What One Word Describes Your Current State of Mind, 2009.

Quem? The New York Times.

O que é? Um visualizador de palavras produzido para refletir as emoções e o humor dos leitores do jornal apoiadores dos candidatos Barack Obama e John MacCain durante a corrida presidencial de 2009. Foi criado a partir da colaboração dos leitores, que enviaram as palavras traduzindo seus sentimentos para o jornal. As palavras são hierarquizadas por tamanho de acordo com a quantidade de pessoas que enviaram uma mesma palavra.

http://www.nytimes.com/interactive/2008/11/04/us/politics/20081104_ELECTION_WORDTRAIN.html

Is It Better to Buy or Rent?, atualizado em 2011

Quem? The New York Times.

O que é? Um infográfico interativo onde o leitor descobre o que pode ser mais vantajoso para ele: alugar ou comprar um determinado imóvel? O infográfico funciona a partir de informações que são digitadas pelo próprio leitor, como o valor mensal de um aluguel, o valor de venda da residência, taxas anuais da propriedade, etc. A ferramenta mostra se é mais vantajoso alugar ou comprar o imóvel em questão a partir de um gráfico, além de mostrar a tendência de valores do imóvel depois de alguns anos.

<http://www.nytimes.com/interactive/business/buy-rent-calculator.html>

Obama's 2011 Budget Proposal: How It's Spent

Quem? The New York Times.

O que é? Gráfico dividido em retângulos cujos tamanhos refletem a quantidade do orçamento do governo norte-americano para 2011 que será gasto em uma determinada categoria, como defesa nacional, saúde, educação, administração pública, etc.

<http://www.nytimes.com/interactive/2010/02/01/us/budget.html>

Investigate your MP's expenses, 2010

Quem? The Guardian.

O que é? Uma página mantida pelo jornal que permite aos leitores investigar e classificar as 458.832 páginas de documentos sobre despesas do governo britânico, disponibilizados por ele, ou conferir o trabalho de análise que já foi feito pelo jornal ou por outros leitores. A página também mantém um registro do que já foi analisado e do que falta ser investigado.

<http://mps-expenses.guardian.co.uk/>

Faces of the Fallen

Quem? The Washington Post.

O que é? Uma página interativa que lista todos os 6.292 soldados norte-americanos que morreram nas operações "Iraq Freedom" e "Enduring Freedom". A página mostra 95 painéis com as fotos dos soldados que faleceram, disponibilizando informações sobre eles e relacionando suas imagens a estatísticas como idade, razão da morte, segmento do exército ao qual pertencia, patente, etc.

<http://apps.washingtonpost.com/national/fallen/>

Casos de dengue no município do Rio em 2011

Quem? Jornal *O Globo*.

O que é? Mapa que mostra a mortalidade de motociclistas por habitante no Brasil, bem como as áreas de concentração de altas taxas de risco em regiões específicas do Brasil.

<http://oglobo.globo.com/infograficos/mapa-da-dengue/>

A epidemia de mortes de motociclistas no Brasil

Quem? Jornal *Estado de S. Paulo*.

O que é? Mapa que mostra os registros de mortes causadas pelo mosquito Aedes aegypti no município do Rio de Janeiro. O mapa mostra a evolução dos casos mês a mês, bem como suas regiões de ocorrência. É possível ainda realizar uma busca das ocorrências por bairro.

<http://www.estadao.com.br/especiais/a-epidemia-de-mortes-de-motociclistas-no-brasil,153378.htm>

8.2 - Entrevista realizada com Alberto Cairo em 25/11/2011, por telefone.

Alberto Cairo é diretor de infografia e multimídia da *Revista Época* e uma das principais referências do jornalismo contemporâneo em infografia e *data-driven journalism*. Entre 2005 e 2010, foi professor na Escola de Jornalismo da Universidade da Carolina do Norte, em Chapel Hill, onde lecionava infografia impressa e online. Entre 2000 e 2005, foi responsável pela infografia online do jornal espanhol *El Mundo*.

- Qual o objetivo da infografia no jornalismo?

Alberto Cairo: A infografia no jornalismo tem os mesmos objetivos que qualquer outro gênero jornalístico: deixar a informação mais clara para os leitores. O que diferencia o recorte entre os gêneros jornalísticos não é tanto os objetivos, mas sim as ferramentas. As ferramentas que a infografia usa estão relacionadas com programação visual, estatística, etc. A infografia é uma combinação de texto com ferramentas visuais.

- Você considera que a infografia é uma reportagem também ou são coisas diferentes?

AC: Depende muito do infográfico que se está fazendo. Alguns infográficos são autônomos e funcionam como reportagens autônomas. Por exemplo, temos na Revista Época uma seção de diagramas que é composta por reportagens feitas a partir de gráficos. Em outros casos a infografia é complemento. Um complemento da informação, um complemento da reportagem escrita, no sentido que ajuda a sustentar as hipóteses da matéria, por meio de providenciar referência visual do que a matéria está contando. Então a infografia tem uma função de complementar ao texto. Mas o texto também pode ser complementar a infografia, dependendo dos casos.

- Quais os desafios para a produção de um bom produto de visualização de dados?

AC: Os desafios batem um pouco com os passos do data-driven journalism. O primeiro desafio é sempre conseguir os dados. Porque os dados nem sempre estão disponíveis, se estão disponíveis pode ser que não estejam no formato certo, às vezes são não contrastáveis pois são de só uma fonte e você tem que confiar nessa fonte. Então há o

desafio de obtenção e tratamento de dados. Agora, na hora de apresentar os dados, você tem vários desafios diferentes dependendo do tipo de gráfico que você esteja criando. O desafio principal é transmitir da forma gráfica uma visão das proporções, das relações, das causalidades, o que não é sempre fácil de obter e nem sempre pode ser conseguido de uma forma intuitiva. Você pode errar e construir o gráfico da forma errada. Por exemplo, o gráfico de bolha: ele às vezes é usado em substituição ao gráfico de barras, o que é errado porque o gráfico de bolha não deixa perceber as proporções de duas magnitudes do mesmo jeito que um gráfico de barras o faz, de forma precisa. Então há o desafio na visualização que o jornalista, ou o design que esteja fazendo o infográfico se pergunte para que este gráfico tem que servir, qual é o objetivo do gráfico. O gráfico no final não é uma ilustração, não é um desenho. É uma ferramenta, uma ferramenta cognitiva que o leitor vai usar para interpretar e analisar os dados. Então, como qualquer outra ferramenta, a forma que essa ferramenta adota deve ser adaptada às funções que essa ferramenta tem. A escolha das formas gráficas não depende do gosto estético do designer, mas sim das funções que esse gráfico tem que ter, das respostas às perguntas que esse gráfico tenta oferecer.

- Qual a diferença de se trabalhar com infografia hoje comparado com o trabalho que era feito há 15 anos, por exemplo?

AC: Há 15 anos, quando comecei a trabalhar com visualização de dados, o trabalho de infografia era todo baseado em ferramentas estáticas. Você tinha um monte de dados e dava forma gráfica a eles em ferramentas como o Illustrator, o Excell, etc. Hoje temos um monte de possibilidade de ferramentas de programação, que estão ficando mais acessíveis para gente que não tem conhecimentos profundos de programação. Por exemplo, o Tableau, que deixa criar visualizações internéticas muito complexas, por meio de um software bem simples de usar. Então nesse tempo a visualização de dados tem avançado primeiro na técnica e, na visualização de impresso, tem avançado muito na parte conceitual. Há 15 anos, a infografia dentro dos jornais era uma área isolada, uma área que olhava para si mesma. Hoje ela olha para fora, para disciplinas que são afins, como a cartografia, para a computação, as novidades que estão sendo criadas para incorporá-las a sua produção diária.

- Um jornalista pode ser um bom repórter e um bom infografista?

AC: A minha ideia hoje em dia é que visualizações simples, e às vezes não tão simples, se você é um bom repórter e tem um pouco de conhecimento conceitual de como são criados os gráficos, mas sem conhecer as ferramentas específicas da infografia como o Illustrator, é possível criar visualizações simples, mas funcionais, com ferramentas como o Tableau. Então, é possível ser um bom repórter e ao mesmo tempo criar bons gráficos? Sim, é possível. Agora, outra coisa é você querer criar visualizações complexas. Aí sim requer alguns especialistas. Mas esse tipo de coisa é criado por equipes, com designers, programadores, etc que ajudam a dar forma a esses gráficos.

- Quais os papéis das bases de dados no jornalismo?

AC: Tem teóricos, como Lev Manovich, que falam que a linguagem do jornalismo moderno é a linguagem das bases de dados. Sem ser tão radical como Manovich, é verdade que as bases de dados hoje são uma importante ferramenta para o jornalismo. O problema é que não há muitos jornalistas que não sabem mexer com bases de dados. Pior ainda, muitos profissionais desconhecem até mesmo as possibilidades que as bases de dados oferecem. Por que você até pode não saber lidar com bases de dados, mas sabendo o que pode ser feito com elas e você contrata alguém que saiba fazer isso. Então esse é um desafio para os jornalistas, e um desafio importante porque hoje em dia, junto com o crescimento das leis de transparência, em vários países, o acesso a grandes quantidades de dados vai ser cada vez mais comum. Então nas redações, será importante contar com gente que saiba mexer e manipular essas bases de dados, essas ferramentas.

- Como você define data-driven journalism?

AC: *Data-driven journalism* não deixa de ser o velho e bom jornalismo de precisão, que é aquele que usa as ferramentas conceituais - estatísticas, softwares - na ciência em geral, mas sobretudo nas ciências sociais. Ele se baseia nas análises de bases de dados, números, pesquisas, etc, para extrair de todas essas ferramentas histórias para contar aos leitores. Então não é nada novo. O que muda é a quantidade, a qualidade e as facilidades

de uso das ferramentas disponíveis. Mas conceitualmente, é idêntico ao bom e velho jornalismo de precisão de Philip Meyer.

- A infografia seria um dos passos de produção desse *data-driven journalism*?

AC: Sim, a visualização, a infografia é um dos passos desse *data-driven journalism*. Aí entra vários fatores: entra a pessoa que apura a informação, a pessoa que trata os dados, a pessoa que cria a base de dados, e nós, que criamos a visualização de dados. Às vezes, não há uma divisão muito clara desses trabalhos. Aqui na Revista Época, por exemplo, eu faço todos os passos, mas são funções que podem ser feitas cumpridas por pessoas diferentes. Nos jornais grandes, como o New York Times, todas as funções estão divididas.

- Você concorda com o alardeamento do *data-driven journalism* como um dos mais promissores campos do jornalismo no século XXI?

AC: Eu sou um pouco cético com as proclamações do tipo "isso vai ser o futuro, isso vai salvar o jornalismo". Acontece que sempre que surge uma novidade, alguma suposta novidade, parece que vai salvar o jornalismo. Como por exemplo o que acontece com o iPad. O iPad vai salvar o jornalismo? Não. Mas uma ferramenta a mais para usar. O mesmo acontece com o *data-driven journalism*. Vai ser a salvação do jornalismo? Não. Vai ser algo que é imprescindível aprender a usar para garantir o futuro do jornalismo. Mas o raciocínio não funciona sem as habilidades tradicionais do jornalismo.

- Você acredita que o movimento do "open data" é uma tendência que tende a se firmar?

AC: Isso sim é uma tendência que eu acho que vai continuar e vai melhorar se o cidadão em geral ficar ciente que isso é importante para ele. E se ele e os governantes tomarem consciência de que os dados que o governo tem não são propriedades do governo, mas sim dos cidadãos. Isso é uma mentalidade que, sobretudo nos países de origem latina, como o Brasil, precisa se firmar ainda. A mentalidade vigente é que se o governo disponibiliza esses dados, ele está fazendo um favor. O que não é correto já que quando você solicita isso, você está pedindo algo que pertence a você, algo a que você

tem direito e algo que é obrigatório, ou deveria ser obrigatório, ao governo fornecer. E não só fornecer esses dados, mas fornecê-los em formatos flexíveis, que é algo que no Brasil precisa avançar um pouco já que os dados são fornecidos normalmente em formato pdf, que são difíceis de lidar. Então é uma tendência que vai continuar? Eu acredito que sim. Mas desde que não pare de se ter uma pressão por essa tendência. Uma pressão da sociedade, de movimentos como o de "open data", de hackers, etc.

- E como você avalia a produção do *data-driven journalism* hoje? É um campo que está engatinhando ou já há produções que realmente já exploram as possibilidades que o *data-driven* oferece?

AC: É um campo que ainda está se desenvolvendo. Tem muitos grupos trabalhando nisso e ele está avançando muito rápido. O *Guardian* está fazendo coisas muito boas. O *New York Times* está fazendo coisas interessantes, o *Washington Post* também. O *Miami Herald* já é tradicional nesse campo, já tendo ganhado um *Pulitzer* por um trabalho de *data-driven journalism*, muito antes que tudo isso fosse tendência. Na Espanha não há muita coisa ainda, mas as primeiras iniciativas já começaram a aparecer. Aqui no Brasil está começando também. É uma coisa incipiente, mas que já está começando. O jornal *O Estado de São Paulo*, a *Revista Época* já dão seus primeiros passos no campo, por exemplo.

- Quais os desafios do *data-driven journalism* nesse início de século XXI, principalmente para que ele possa se disseminar nas redações jornalísticas como prática?

AC: A maior dificuldade é o acesso aos dados. Esse é o primeiro desafio que temos e isso pode ser resolvido com leis de acesso público aos dados que garanta os direitos e as obrigações dos governos para liberarem os dados nos formatos certos. O segundo grande desafio - um desafio muito importante - é que você não consegue fazer nada que precise de formação especializada se você não tem as pessoas certas na redação. Isso é algo que acontece muito nas empresas de comunicação, sobretudo as de jornalismo: você não tem pessoas nas redações com "know how", que saibam mexer com números. Isso é totalmente escandaloso. Estou falando dos grandes veículos de comunicação, quanto os médios, porque nas faculdades de jornalismo é incompreensível que não

ensinem fundamentos de estatística, o que é uma coisa totalmente fora da realidade. A estatística tem que ser obrigatória na formação de jornalistas. Então o que falta muito nas redações, é formação teórica e técnica neste tipo de campo, o que é um grande desafio. E depois a falta de formação técnica dos próprios designers. Já trabalhei em departamento gráficos de jornais em que os profissionais não tem cultura jornalística e nem de estatística.

