

MATEUS DE ORNELLAS RODRIGUES

# **BASES DE DADOS PÚBLICAS NA PRODUÇÃO DO JORNALISMO *ONLINE***

Monografia de Conclusão de Curso de Especialização em Jornalismo Multimídia, exigência parcial para obtenção do título de Especialista em Jornalismo Multimídia pela Pontifícia Universidade Católica de São Paulo.

**Orientador: Prof. André Deak**

Coordenadoria Geral de Especialização, Aperfeiçoamento e  
Extensão da PUC-SP

Pontifícia Universidade Católica de São Paulo

São Paulo, dezembro de 2010.

## **AGRADECIMENTOS**

À minha família, pelo amor e apoio incondicional que sempre recebi. Nada do que eu pudesse escrever seria capaz de expressar toda a gratidão que eu tenho por todos esses que constituem base de qualquer eventual sucesso que eu possa vir a alcançar em minha vida.

Ao meu orientador, Prof. André Deak, pela dedicação ao trabalho que nos propusemos realizar. Tê-lo como mestre permitiu que eu me desenvolvesse intelectualmente e ampliou minha paixão à minha carreira e meus estudos.

À Bárbara, por sempre tolerar minha falta de tempo e meu cansaço com um sorriso. Sua presença ao meu lado tornou este caminho um pouco menos árduo.

A todos os meus amigos, suporte essencial para os meus dias se tornarem realmente completos. Destes, dois merecem uma menção especial. Nityananda Portellada, meus agradecimentos pelo auxílio para que eu compreendesse algumas questões de natureza jurídica que permearam partes desta pesquisa; e Ricardo Cesar Bernardo Souza, pelo inestimável auxílio como cinegrafista para que eu pudesse registrar algumas das entrevistas realizadas em áudio e vídeo.

## RESUMO

Esta pesquisa tem como objetivo compreender como se dá o uso de bases de dados informacionais públicas no jornalismo produzido para meios digitais de informação e comunicação. A partir de entrevistas realizadas com profissionais que atuam diretamente com o jornalismo produzido a partir de bases de dados (*database journalism* ou *data-driven journalism*) e militantes da transparência governamental, buscou-se estabelecer a relação entre a disponibilização de informação pública e a produção de informação jornalística a partir do intercruzamento, análise das informações e interpretação para a construção de uma narrativa jornalística. Pretendeu-se também debater sobre as possibilidades de contribuição destes métodos para o desenvolvimento do jornalismo que é produzido para as mídias digitais.

**PALAVRAS-CHAVE:** jornalismo online; bases de dados; transparência pública; jornalismo a partir de bases de dados

## **ABSTRACT**

This research aims to understand the use of public informational databases in journalism produced for digital media of information and communication. From interviews with professionals who work directly with the journalism produced from databases (database data-driven journalism or journalism) and militants of government transparency, we sought to establish the relationship between the availability of public information and production of journalistic information from interbreeding, analysis and interpretation of information for the construction of a journalistic narrative. We also intend to discuss the possible contribution of these methods to the development of journalism that is produced for digital media.

**KEYWORDS:** online journalism, databases, public transparency; data-driven journalism

## SUMÁRIO

Introdução 6

1. Jornalismo Online a partir de Bases de Dados: definição e origens 12
  - 1.1 O que são bases de dados? 12
  - 1.2 O papel das bases de dados nas tecnologias digitais de informação e comunicação (TIC's) 14
  - 1.3 Desenvolvimentos do uso de bases de dados no jornalismo online 16
    - 1.3.1 O jornalismo de precisão e a Reportagem com Auxílio de Computador (RAC) 16
    - 1.3.2 Da RAC ao jornalismo online em Base de Dados 20
    - 1.3.3 Jornalismo Online feito a partir de bases de dados 21
2. Bases de dados públicos e informação pública 30
  - 2.1 - Bases de dados públicas e transparência 30
  - 2.2 Acesso e possibilidade de uso dos dados públicos 36
3. Práticas e rumos do uso de bases de dados no jornalismo online 41
  - 3.1 A estrutura das reportagens digitais baseadas em dados públicos 41

Conclusão 48

Metodologia 52

Bibliografia 57

## INTRODUÇÃO

O surgimento e desenvolvimento das tecnologias digitais de informação e comunicação é ponto comum de uma série de importantes mudanças na sociedade. CASTELLS (1999) define essa “revolução da tecnologia da informação” como “no mínimo, um evento histórico da mesma importância da Revolução Industrial do século XVIII, induzindo um padrão de descontinuidade nas bases materiais da economia, sociedade e cultura” (p. 68). Segundo o autor, diferentemente do que aconteceu em outros períodos históricos nos quais as informações e conhecimentos serviam como combustível para o desenvolvimento das tecnologias que geravam a revolução, o desenvolvimento de novos conhecimentos e informações servem para gerar ou aprimorar os próprios dispositivos de comunicação, informação e geração de conhecimento, num processo de retroalimentação de uso e invenção (CASTELLS, *idem*).

A *internet* talvez seja a mais representativa em relação à importância dessas tecnologias nessa sociedade em rede. O levantamento realizado pelo CGI-Br<sup>1</sup> demonstra que 45% dos brasileiros já utilizou a internet e, desse total, 39% a havia utilizado nos últimos três meses. Desse total, 90% utilizou a internet para fins de comunicação (envio de e-mail, *chats* e conversas através de programas de mensagens instantâneas). Outra informação relevante que foi obtida através da pesquisa foi a de que, do total de pessoas que utilizaram a internet nos últimos três meses, 86% a utilizaram para o lazer. Destes, 40% a utilizaram para ouvir rádios *online* e 44% leram jornais e/ou revistas através da internet. De acordo com esta pesquisa, o ato de se informar ocupa uma parcela importante do tempo que as pessoas dedicam à *internet*.

A importância que os veículos informativos que atuam na Rede também vêm crescendo. Um estudo publicado pelo *Pew Research Center for the People & the Press* em dezembro de 2008<sup>2</sup> mostra que 40% dos entrevistados pela pesquisa

---

<sup>1</sup>De acordo com informações do relatório “Pesquisa sobre Uso das Tecnologias da Informação e Comunicação no Brasil 2009”.

<sup>2</sup>Disponível em <http://people-press.org/reports/pdf/479.pdf>. Acessado em 03 set 2010. A pesquisa foi feita com uma amostragem de 1,000 entrevistas feitas por telefone com pessoas de mais de 18 anos nos Estados Unidos. O percentual de erro é de 3,5 pontos, podendo ser maior quando tratando de opiniões de subgrupos.

confiam mais na internet como meio de obtenção de notícias de relevância nacional e internacional em setembro de 2007, superando o número de pessoas que citaram o jornal como seu principal meio de obtenção de notícias sobre este assunto (35%). A pesquisa também cita que, entre as pessoas com menos de 30 anos entrevistadas sobre o assunto, 59% afirmam que obtém a maior parte de suas notícias sobre assuntos nacionais e internacionais através da internet. Este percentual é igual ao de pessoas que declaram que a televisão é seu principal meio de acesso a este tipo de notícia.

Sem a intenção de entrar em discussões sobre o destino do jornal ou dos outros meios analógicos ou mesmo a suposta “competição” de audiência destes meios com a *internet*, os dados dessa pesquisa são importantes já que revelam que o jornalismo produzido para meios digitais ocupa um espaço cada vez mais importante na vida das pessoas e, por consequência, no panorama geral da atividade.

### **Definição para o jornalismo feito para os meios digitais de comunicação**

Antes de prosseguir, é necessário esclarecer alguns pontos da terminologia e de conceitos que foram utilizados nesta pesquisa para conceituar a prática do jornalismo nos ambientes proporcionados pelas novas tecnologias de informação e comunicação. Não é, de forma alguma, intenção deste trabalho esgotar as discussões que estão sendo realizadas quanto as melhores definições para estas atividades. Porém, considera-se importante esclarecer, mesmo que de forma sucinta, os critérios para a escolha do conjunto de termos utilizados nesta pesquisa, visando contribuir para a precisão do trabalho científico realizado e a melhor compreensão do leitor deste trabalho.

BARBOSA (2007a) utiliza a definição *jornalismo digital* como sinônimo de *jornalismo online*, *webjornalismo*, *ciberjornalismo* e *jornalismo contextualizado*. Segundo a pesquisadora, o jornalismo digital é definido como aquele “(...) feito no âmbito da internet e que pressupõe a coleta, produção, publicação e disseminação de conteúdos através da web e também de outros dispositivos como celulares,

PDA's, etc (...)” (BARBOSA, *idem*: p.127-128).

FERRARI (2003) considera que o termo *jornalismo digital* “compreende todos os noticiários, *sites* e produtos que nasceram diretamente na *web*” (p. 41). Ela difere essa definição da de *jornalismo online*, que seria aquele realizado através da transposição das mídias, ou seja, da conversão do conteúdo produzido para as mídias tradicionais para a *internet*, como nos *sites* de jornais ou revistas que não possuem um núcleo dedicado à produção de conteúdo específico e se servem da conversão daquilo que é publicado no papel.

DEAK (2011) decidiu-se pelo conceito do termo *ciberjornalismo*, com preferência de uso da expressão “jornalismo na *internet*” para beneficiar uma compreensão mais ampla. “Tanto um quanto o outro são entendidos (...) como o jornalismo produzido *para* o ciberespaço, e distribuído nele, a ser recebido em qualquer equipamento que o acesse” (DEAK, *idem*, p. 22).

A partir dessas concepções, decidiu-se nessa pesquisa adotar o termo *jornalismo online* para definir o jornalismo produzido para o ambiente de redes digitais de comunicação e que se utiliza desse ambiente como ferramenta para seu desenvolvimento. Com este esclarecimento, espera-se que a terminologia seja clara o suficiente para evitar qualquer erro de compreensão do texto, sem reduzir a discussão sobre essa nomenclatura ou dedicar um espaço maior a tal questão, que não é um dos objetivos deste trabalho.

## **O surgimento do jornalismo online e do uso de bases de dados informatizadas**

Embora praticamente nenhuma atividade ou ramo de conhecimento tenha permanecido inalterado após o desenvolvimento das tecnologias de informação e comunicação, o jornalismo provavelmente figura entre um dos mais modificados por esta torrente de desenvolvimentos que tem suas origens nas experiências de matemáticos e engenheiros, como Charles Babbage e Ada Lovelace, no fim do século XIX e que se acentuou após o surgimento da informática a partir da segunda metade do século XX. O surgimento do jornalismo online se dá em paralelo com a

disseminação da *Internet* em sua fase comercial. Embora as primeiras tentativas registradas tenham acontecido em meados da década de 80, em primeiras experiências em videotexto, o crescimento maciço dos esforços para se estabelecer uma produção jornalística para a internet só acontece após a criação da World Wide Web, em 1989.

A criação da WWW foi a principal responsável pela ampliação da utilização da rede, já que até aquele momento sua utilização era limitada aos meios militares e acadêmicos principalmente pela dificuldade de sua compreensão e utilização para aqueles que não eram especialistas no assunto. O desenvolvimento da WWW pelo Centre Européen pour Recherche Nuclear (CERN) em 1990 permitiu que diferentes computadores pudessem exibir as páginas criadas de forma próxima da uniformidade graças ao desenvolvimento dos protocolos de conexão TCP/IP, que permitia a interconexão entre diferentes microcomputadores, e do protocolo de transferência de hipertexto HTTP, que orientava a comunicação entre os programas navegadores (ou *browsers*) e os servidores, conforme explica CASTELLS (*idem*).

Graças às possibilidades geradas por uma interface que dispunha dos hiperlinks e recursos multimídias trazidas pela WWW, o número de jornais que ingressaram na internet aumentou de forma rápida e notável. Segundo André Manta<sup>3</sup>, apenas vinte jornais estavam presentes na internet em 1993. Em 1997, chegavam a dois mil jornais e mais de quatro mil revistas digitais em todo o mundo.

No Brasil, o Jornal do Brasil e o Estado de S. Paulo estabeleceram seus respectivos *websites* no começo de 1995, sendo os primeiros do país a terem conteúdo publicado na internet, de acordo com BORGES (2009). De acordo com o autor, a equipe inicial do JB era composta por apenas três pessoas e apenas reproduzia algumas notícias da edição impressa do jornal. A partir do ano seguinte, passou-se a reproduzir todo o conteúdo do jornal impresso na internet.

A tarefa de determinar cronologicamente a utilização de bases de dados no jornalismo brasileiro é uma tarefa árdua, devido aos poucos registros mantidos sobre

---

<sup>3</sup>De acordo com informações disponíveis em <http://www.facom.ufba.br/pesq/cyber/manta/Guia/cap02.html#Nota2>. Acessado em 18 set 2010.

o assunto pelas empresas jornalísticas e a falta de pesquisas acadêmicas que deem conta desta periodização. Inicialmente, as bases de dados digitais foram utilizadas para a manutenção dos arquivos de texto e imagens do jornal, para serem acessados pelos jornalistas da empresa, antes mesmo da implantação de uma versão do jornal para a internet. Este processo ocorreu durante a década de 1980 e 1990 (BARBOSA, 2007b). Na Folha de S. Paulo, este processo se iniciou em 1993, com a digitalização dos arquivos do jornal para acesso interno através dos terminais da empresa<sup>4</sup>. O Estado de S. Paulo realizou um processo similar. De acordo com a classificação proposta por MACHADO (2006), essa utilização das bases de dados serve com o intuito de preservar a memória dos veículos jornalísticos. A informatização dos arquivos do jornal permite o acesso rápido e simultâneo por vários profissionais aos conteúdos armazenados.

A ampliação da utilização das bases de dados para o jornalismo online seguiu as possibilidades abertas pelo aprimoramento tecnológico crescente da internet. Ainda segundo BARBOSA (*idem*), esse tipo de utilização de bases de dados para a distribuição de produtos jornalísticos começou nos primeiros anos da década de 1990, com os esforços para a criação de *websites* jornalísticos mais dinâmicos. Esta transição seguiu o ritmo permitido pelos desenvolvimentos no campo da informática, com a criação de linguagens para a *web* que possibilitaram a criação de *sites* que mantivessem uma espécie de infraestrutura fixa e que fossem preenchidos com informações advindas de bases de dados atualizáveis e determinadas a ocupar aquele espaço na estrutura. Assim, tais *sites* permitiam que partes fossem atualizadas de modo mais rápido e fácil sem a necessidade de alteração em toda a sua programação para a inserção da nova informação desejada.

Nesta pesquisa, focaremos em conteúdos jornalísticos produzidos a partir de bases de dados internas – produzidas pelo jornalista ou pela entidade jornalística a qual ele está associado a partir de informações apuradas – ou externas – a partir de bases de dados já estruturadas externamente e abertas à utilização – organizadas em uma determinada interface informativa para o acesso ao leitor final. Dessa forma, pretende-se restringir a pesquisa a produções que utilizem as bases de dados como

---

<sup>4</sup>De acordo com informação disponível em [http://www1.folha.uol.com.br/folha/circulo/historia\\_folha.htm](http://www1.folha.uol.com.br/folha/circulo/historia_folha.htm).  
Acessado em 05 10 2010.

recurso para a produção de informação jornalística específica e não como suporte tecnológico para a distribuição de outros tipos de conteúdos, como é o caso de *websites* jornalísticos que são montados a partir de bases de dados que são alimentadas com diversos tipos de conteúdos – galerias de imagens, textos e vídeos.

No primeiro capítulo, será tratado o uso de bases de dados pelo jornalismo *online*. Traçaremos sua origem nas primeiras utilizações de computadores nas redações jornalísticas a partir de meados de 1980. Passaremos pelo desenvolvimento do jornalismo de precisão até o desenvolvimento da Reportagem Assistida por Computador (RAC), precursora do que hoje podemos denominar como Jornalismo a partir de Bases de Dados. Finalmente, analisaremos alguns exemplos do uso de bases de dados públicas no Jornalismo a partir de Bases de Dados.

O segundo capítulo foi dedicado a buscar compreender historicamente como os dados públicos começaram a ser disponibilizados para a sociedade e como se dá a luta por uma ampliação do acesso a tais dados, além de tentar estabelecer sua relação com os conceitos de governo aberto e transparência governamental. Finalmente, buscou-se entender a relação entre o acesso a dados públicos e o desenvolvimento do jornalismo *online*.

Finalmente, o último capítulo foi dedicado a compreender as características mais importantes desse jornalismo que se desenvolve a partir de bases de dados públicas. Como ele é construído, que tipo de conhecimento é exigido do profissional que deseja utilizar esta técnica e que tipo de contribuição esta prática pode dar para o desenvolvimento do jornalismo produzido para os meios digitais de informação e comunicação.

## 1. Jornalismo Online a partir de Bases de Dados: definição e origens

### 1.1 O que são bases de dados?

É importante que seja definido o conceito de bases de dados<sup>5</sup> que será utilizado na pesquisa aqui descrita. O conceito, que pode ser considerado sinônimo de *banco de dados*, conforme o uso mais habitual que é feito hoje, tem sua origem na Ciência da Computação. ELMASRI e NAVATHE (2011) definem banco de dados como:

“[...] uma coleção de dados relacionados. Com dados, queremos dizer fatos conhecidos que podem ser registrados e possuem significado implícito. Por exemplo, considere os nomes, números de telefone e endereços das pessoas que você conhece. Você pode ter registrado esses dados em uma agenda ou, talvez, os tenha armazenado em um disco rígido, usando um computador pessoal e um software como *Microsoft Access* ou *Excel*. Essa coleção de dados relacionados, com um significado implícito, é um banco de dados” (p. 3).

Como os próprios autores admitem, essa definição é muito ampla, pois admite que qualquer coleção de itens relacionada por qualquer tipo de critério atende aos requisitos. Eles propõem alguns critérios adicionais para o que eles chamam de “uso comum” - ou melhor, entendimento comum – do que é um banco de dados. Para ELMASRI e NAVATHE (*idem*):

“- um banco de dados representa algum aspecto do mundo real, às vezes chamado de **mini-mundo** ou de **universo de discurso (UoD – Universe of Discourse)**. As mudanças no minimundo (*sic*) são refletidas no banco de dados.

– Um banco de dados é uma coleção logicamente coerente de dados com algum significado inerente. Uma variedade aleatória de dados não pode ser corretamente chamada de banco de dados.

---

<sup>5</sup>Os termos *bases de dados* e *banco de dados* são sinônimos. Optou-se, nessa pesquisa, por utilizar o primeiro termo, seguindo a nomenclatura proposta por BARBOSA (2007a, 2007b) e MACHADO (2006), entre outros. Pretende-se que isso ajude na solidificação do termo, contribuindo para a precisão de sua aplicação e entendimento como conceito, além de isso enfatizar o papel dessas informações organizadas como base a partir da qual se obtém o resultado desejado, como a informação no caso do jornalismo digital.

– Um banco de dados é projetado, construído e populado com dados para uma finalidade específica. Ele possui um grupo definido de usuários e algumas aplicações previamente concebidas nas quais esses usuários estão interessados.” (p. 3)

Outros pesquisadores que também se debruçaram sobre a questão do uso de bases de dados também recorreram à Ciência da Computação para defini-las. MACHADO (2006) utiliza a definição proposta por GUIMARÃES (*apud* MACHADO, *idem*) e define a base de dados como:

“uma coleção de dados ou informações relacionados entre si, que representam aspectos de um conjunto de objetos com significado próprio e que desejamos armazenar para o futuro” (p. 16)

Para MANOVICH (2001),

“*base de dados* pode ser definida como um conjunto estruturado de dados. Os dados armazenados em uma base de dados são organizados para que sejam buscados e recuperados rapidamente por um computador e, portanto, podem ser tudo menos uma mera coleção de itens”.

A partir dos pontos comuns entre os conceitos propostos por esses autores, decidiu-se estabelecer uma definição breve e sucinta que permita identificar as bases de dados que serão objetos da pesquisa aqui apresentadas. Dessa forma, definem-se *bases de dados* como *conjuntos de informações de determinado aspecto da realidade reunidas e estruturadas de maneira a que os elementos que a compõem possuam ou produzam algum sentido quando relacionados*.

O surgimento das bases de dados de forma alguma está relacionado com o surgimento ou desenvolvimento da microinformática. A capacidade de contabilizar e interpretar informações são necessidades essenciais para o desenvolvimento de diversas atividades. O registro de colheitas, a contabilização e localização de itens produzidos, os censos populacionais; todos esses registros atendem aos requisitos propostos e devem, com razão, ser consideradas bases de dados. Da mesma forma, é necessário considerar que o advento da informática ampliou a capacidade desses

registros serem armazenados, ampliou a velocidade de processamento e cruzamento dos dados, potencializando dessa forma seus usos e permitindo resultados que antes seriam quase impossíveis de ser obtidos ou que exigiriam um esforço e mão de obra muito grande para serem alcançados.

É importante notar que todos os autores relacionados fazem questão de ressaltar que o fato de ser uma coleção de itens é menos importante para definir o conjunto como uma base de dados do que a importância de que tais itens possuam determinada relação de sentido entre si, ou seja, sejam capazes de produzir informação através de seus relacionamentos. Uma base de dados é, portanto, uma forma de organizar informações a fim de construir uma representação de determinada parcela da realidade a partir de seu mini-mundo, conforme o jargão da informática. A construção e utilização de bases de dados, em todas as atividades nas quais elas são aplicadas, servem para auxiliar no processo de interpretação daquele pedaço de realidade que foi registrado. A partir do intercruzamento sucessivo das informações, é possível estabelecer os resultados que não são óbvios a partir da observação simples das informações armazenadas.

## **1.2 O papel das bases de dados nas tecnologias digitais de informação e comunicação (TIC's)**

Uma breve observação da sociedade contemporânea é suficiente para atestar a importância da organização e utilização de bases de dados combinadas às tecnologias digitais de informação e comunicação para a realização de diversas funções econômicas e sociais. “É correto afirmar que os bancos de dados desempenham um papel crítico em quase todas as áreas em que os computadores são usados, incluindo negócios, comércio eletrônico, engenharia, medicina, genética, direito, educação e biblioteconomia” (ELMASRI e NAVATHE, *ibidem*, p. 2).

A construção e organização de bases de dados cada vez maiores vem permitindo que empresas administrem grandes estoques com uma velocidade muito grande, adequando seu fornecimento a demandas cada vez mais específicas graças às possibilidades de identificar rapidamente a melhor forma para atendê-las,

diminuindo custos de armazenamento e transporte. Para os governos, bases de dados cada vez mais sofisticadas permitem que necessidades específicas de determinados setores da comunidade ou zonas da região administrada sejam visualizadas. Além disso, as bases de dados permitem um controle social maior, com a identificação de indivíduos através do seu uso integrado a outras tecnologias como as de reconhecimento facial ou de caracteres (como os empregados para a identificação de placas de veículos, por exemplo). Setores completamente díspares também se beneficiam de modo crucial do constante desenvolvimento das bases de dados, desde o mercado financeiro à administração hospitalar, passando pela agricultura e pecuária e até a educação. Na verdade, é difícil identificar uma atividade humana atual que não use ou se beneficie de alguma forma de base de dados.

MANOVICH (*idem*) afirma que a base de dados pode ser chamada de “nova forma simbólica da era do computador” (p. 219). Segundo o autor, a maioria dos objetos das novas mídias – conforme ele denomina as criações, tanto aquelas de cunho artístico quanto as com finalidades objetivas produzidas para os meios digitais de comunicação – se organiza na forma de base de dados. Ele cita as enciclopédias multimídias distribuídas em CD-ROM que não são lidas de forma linear, mas que são exploradas através do acesso a qualquer ponto do conteúdo. O usuário não precisa seguir uma ordem pré-determinada; todos os itens estão virtualmente dispostos de forma a serem acessados com igual facilidade.

A própria organização da internet deve ser interpretada como uma gigantesca base de dados organizada em forma de rede. Cada página, cada *hiperlink*, conduz o leitor para outra organização de dados – no caso, outro conjunto de dados organizados numa página (texto, imagens, sons). Sentidos completamente diferentes podem ser obtidos a partir dos mesmos conteúdos, a partir da interligação que se faz entre eles. A forma de relacionamento entre os *hiperlinks* apresentados, ou a partes específicas de determinada página, podem alterar até mesmo o sentido originalmente desejado pelo criador daquele material. Pensar a *internet* como base de dados ressalta a importância da relação entre os dados estabelecidos e a significação que é obtida a partir de suas conexões.

### 1.3 Desenvolvimentos do uso de bases de dados no jornalismo online

O uso de bases de dados para a produção de conteúdo jornalístico não se inicia a partir das obras produzidas especificamente para a *internet* e outros dispositivos digitais e nem é uma característica exclusiva do que é produzido para esse meio. Da mesma forma, o cruzamento das novas tecnologias com a produção jornalística criou um ambiente propício para que essa combinação tivesse seu potencial amplificado. A seguir, busca-se traçar um breve histórico do desenvolvimento e utilização de base de dados por parte do jornalismo online.

#### 1.3.1 O jornalismo de precisão e a Reportagem com Auxílio de Computador (RAC)

Os computadores adentraram as redações dos jornais a partir da década de 60, seguindo a tendência dos setores empresariais e industriais de se utilizarem da nova tecnologia para aperfeiçoar os processos de produção, realizando as tarefas em menor tempo e com mais precisão (BARBOSA, 2007b, p. 98-99). Inicialmente, os computadores serviam basicamente para agilizar os processos de produção e edição de notícias. Posteriormente, os computadores também foram aplicados na composição das páginas dos jornais e revistas.

As mudanças proporcionadas pela informática nas redações brasileiras também começaram a surgir na década de 1960, embora tenham levado muito mais tempo para serem aplicadas de forma efetiva e causarem alguma mudança sensível na rotina de trabalho desses ambientes.

“No Brasil, a *Folha de S. Paulo* foi o primeiro a introduzir terminais de computadores em suas redações, projeto que vinha sendo elaborado desde 1968 e foi efetivado em 1983. Como a *Folha*, *O Globo*, *O Diário Catarinense*, *A Tribuna*, *Zero Hora*, *O Estado de S. Paulo* também se adaptaram à nova era do jornalismo, assim como revistas, a exemplo da *Veja*” (BARBOSA, *idem*, p. 100).

SOARES (2011), ao falar de sua experiência como *office-boy* do jornal *Correio*

do Povo de Porto Alegre - RS, exemplifica como a chegada das novas tecnologias rapidamente modificaram os processos de produção das edições do jornal:

“Quando eu entrei no Correio do Povo, para colocar uma foto na página, foto de arquivo [...]; o editor pegava, anotava num papelzinho, fazia uma requisição de pasta. Dizia: 'Ah, eu quero a pasta com as fotos do Frank Sinatra'. [...] Na sala do arquivo, os caras iam buscar a pasta do Frank Sinatra. Me davam aquele catatau de pasta e eu levava lá pro editor. O editor escolhia qual era a foto, mandava pro diagramador.

O diagramador colava um papelzinho, chamado '*porrinha*', colava ali na foto dando o tamanho que ele queria a foto. Aí, Seu Marcelo pegava de novo a foto com a '*porrinha*', levava lá no fundo do corredor, que era a fotomecânica. [...] Uma sala inteira só para mudar o tamanho das fotos. Tinha uma câmara do tamanho de um elevador [...] onde os caras botavam a foto, eles davam o tamanho que eles queriam. Ampliavam, reduziavam no tamanho que o editor quisesse. Aí eu ia lá buscar a foto de novo pra levar no *paste-up*. *Paste-up* era uma placa de acrílico grandona onde eles colavam o modelo da página riscada. [...] Isso é 95, isso faz 15 anos, 16 [anos] agora. Aí, eles tinham o modelo da página, esse cara dizia “Aqui vai o título, aqui vai a matéria, aqui vai a foto e tal”. E eles iam colando, elemento por elemento da página naquele modelo. [...] Depois, no final da noite, seria levado de volta para fotografar, fazer o fotolito. No ano seguinte, já mudou isso aí. As fotos passaram a ser digitalizadas. A gente tinha que buscar a foto no arquivo ainda, mas para levar para outro departamento, para digitalizar e tal. Depois mudou o fotolito; foi mudando muito rápido, de um ano pro outro, sabe? Hoje em dia, o que tu faz com um cara só, pra botar uma foto na página, precisava de dez. E eu peguei essas mudanças todas ali, observando como [*office-*]boy.” (p. 4)

A próxima mudança importante relacionada à chegada dos computadores nas redações foi a geração da Reportagem com Auxílio de Computador (RAC), considerada ponto de origem para o que hoje é chamado de Jornalismo de Base de Dados de acordo com jornalistas e pesquisadores como SOARES (*idem*) e GOULARD (2010). A RAC consiste em técnicas para a utilização de computadores na busca e filtragem de informações, permitindo a análise de grandes quantidades de informação em menos tempo, ajudando na apuração e produção de pautas<sup>6</sup>.

---

<sup>6</sup>Conforme definição proposta em “Curso de RAC - Reportagem com Auxílio do Computador ou Como transformar o computador no melhor amigo do jornalista”. Disponível em [http://www.arquivoz.com.br/arquivo/Curso\\_CAR\\_Abraji\\_rio.pdf](http://www.arquivoz.com.br/arquivo/Curso_CAR_Abraji_rio.pdf). Acessado em: 21 jun 2011.

O desenvolvimento da RAC se deu em paralelo à crescente utilização de computadores nos negócios e na indústria a partir da década de 1950. A primeira aplicação da RAC foi feita nas eleições presidenciais estadunidenses de 1952. Walter Cronkite, correspondente da *CBS News* em Washington DC, utilizou um computador e fórmulas preparadas por programadores para prever o resultado da disputa entre Dwight Eisenhower e Adlai Stevenson. Ao contrário da disputa acirrada que era esperada, os resultados obtidos por Cronkite indicavam que Eisenhower venceria com uma larga vantagem. O resultado era tão surpreendente que os responsáveis na CBS adiaram por horas antes de finalmente divulgar esse resultado, que se mostrou correto ao fim da apuração. (COX, 2000: p. 6)

O amadurecimento da técnica se deu na década de 60, a partir do trabalho daquele que pode ser considerado o principal nome da RAC: Philip Meyer. Ele atuou na cobertura dos tumultos acontecidos na cidade de Detroit em julho de 1967, quando uma batida policial num bar sem licença gerou uma série de saques, incêndios e tiroteios num período de seis dias de revolta. No total, estimou-se que 43 pessoas morreram, centenas ficaram feridas e mais de 7.000 foram presas ao fim dos tumultos.

As primeiras teorias formuladas para tentar entender quem eram as pessoas e porque o tumulto começou supunham que eram indivíduos da parte mais baixa da sociedade ou recém-chegados do sul que não haviam se adaptado e agora descontavam suas frustrações na cidade.<sup>7</sup> O trabalho de Meyer fez com que essas ideias caíssem por terra. Ele utilizou um computador para analisar as informações obtidas por uma pesquisa realizada entre os afro-americanos que participaram das revoltas.

“A análise revelou que, ao contrário da hipótese aceita, pessoas que cursaram a faculdade tinham a mesma probabilidade de participar dos confrontos que aquelas que haviam abandonado a escola” (COX, *idem*: p. 7)

Essa experiência pioneira utilizou ferramentas que, até então, eram

---

<sup>7</sup>De acordo com informações obtidas no artigo “Revealing the roots of a riot”. Disponível em <http://www.sampler.isr.umich.edu/2011/featured/revealing-the-roots-of-a-riot/>. Acessado em: 22 jun 2011.

inexistentes na prática jornalística: a aplicação de técnicas e métodos de pesquisa característicos das ciências sociais, como o uso de questionários na apuração dos distúrbios em Detroit; e o uso do computador como uma ferramenta para auxiliar o processo de interpretação de informações obtidas ao invés de apenas limitado ao papel de processamento e edição de textos.

Após a reportagem, que rendeu um prêmio *Pulitzer* à equipe do jornal *Detroit Free Press*<sup>8</sup>, Philip Meyer começa a escrever um livro sobre o uso de pesquisas estatísticas na produção de jornalismo. Em 1973, o livro é publicado sob o título de *Precision Journalism*. Na obra, o autor declara que a sociedade está se tornando cada vez mais complexa e que, com isso, é necessário que o jornalismo repense a forma com que atua.

“Houve um tempo quando tudo o que você precisava era dedicação à verdade, abundância de energia e algum talento para escrever. Você ainda precisa dessas coisas, mas elas não são mais suficientes. O mundo se tornou tão complicado, o crescimento de informação disponível tão explosiva, que o jornalista precisa ser um filtro, bem como um transmissor; um organizador e interpretador, bem como alguém que obtém e entrega fatos. Além de saber como produzir informação para o impresso, online ou para o ar, ele ou ela também precisa saber como chegar à cabeça do receptor. Em resumo, um jornalista precisa ser um administrador de banco de dados, um processador de dados e um analista de dados”. (MEYER, 2002, p. 1)

Meyer compreendeu que a matéria com a qual o jornalismo trabalha é a informação, não o texto ou outro suporte. (SOARES, *ibidem*, p. 6) Também percebeu que, ao trabalhar com os dados primários ao invés de apenas se fixar em metodologias que podem ser consideradas mais “clássicas” da atividade, como entrevistar pessoas, o jornalista amplia sua capacidade de produzir informação de qualidade ou de interesse de seu público.

Em 1968, Clarence Jones do jornal *The Miami Herald* produziu a reportagem “*A Scientific Look at Dade Crime*”<sup>9</sup>, contratando estudantes de direito para inserir os

---

<sup>8</sup>De acordo com informações disponíveis no *site* do Prêmio *Pulitzer*. Disponível em <http://www.pulitzer.org/awards/1968>. Acessado em: 25 jun 2011.

<sup>9</sup>“Um olhar científico sobre o crime em Dade”, em tradução para o português.

registros do tribunal do condado de Dade e posteriormente analisando os 13,000 cartões perfurados resultantes para definir padrões do sistema de justiça criminal (COX, *idem*, p. 9). A reportagem, além de importante para o desenvolvimento para a RAC, pode ser considerada a primeira a utilizar bases de dados públicas através dessa técnica.

No Brasil, o jornalista Marcelo Soares é um dos pioneiros no desenvolvimento de reportagens produzidas com auxílio de computadores. Ele estuda o tema desde 1997, quando ainda era estudante de jornalismo na Universidade Federal do Rio Grande do Sul. Em 2002, Marcelo ajudou na fundação da Associação Brasileira de Jornalismo Investigativo (ABRAJI), entidade da qual foi gerente e instrutor de RAC. A ABRAJI tem um papel importante na disseminação da prática da RAC, ministrando cursos sobre o assunto para diversos veículos de imprensa por todo o país e organizando seminários e congressos sobre o assunto.

### **1.3.2 – Da RAC ao jornalismo online em Base de Dados**

Embora se possa afirmar com razoável segurança que as origens do que determinamos como “jornalismo online em Base de Dados” estejam ligadas à RAC, a tarefa de chegar a um ponto de ruptura, o exato momento onde a prática deixou de ser a primeira para se transformar na última é mais complexa.

Em princípio, parece razoável considerar que, apesar de ser possível identificar seu surgimento num período posterior ao da RAC, o que denominamos como jornalismo a partir de base de dados não é um desenvolvimento ou uma consequência. Na verdade, os instrumentos proporcionados pela RAC podem ser usados na construção de reportagens a partir de bases de dados.

De fato, a utilização cada vez mais disseminada de computadores nas mais diversas tarefas e nos mais diversos pontos da produção jornalística faz surgir a dúvida se a utilização do termo Reportagem Assistida por Computador como prática ainda faz algum sentido. SOARES (2010) argumenta que todo produto jornalístico é produzido a partir de computadores. Assim sendo, toda reportagem seria uma RAC.

Portanto, parece mais razoável concluir que não há um ponto ruptura distinguível ou de efetiva importância entre a prática da RAC e a do jornalismo online em Base de Dados. A utilização mais frequente e com resultados mais expressivos das bases de dados para a produção do jornalismo seria uma espécie de especialização dos conceitos e técnicas da RAC somados a maior oferta de dados públicos e privados das mais variadas espécies, como aqueles produzidos pelos próprios jornalistas a partir da reunião dos resultados de suas apurações diárias, por exemplo.

### 1.3.3 – Jornalismo Online feito a partir de bases de dados

O uso de bases de dados no jornalismo online é uma prática consideravelmente recente dentro do universo do jornalismo e ainda realizada por poucos profissionais. Dessa forma, tentar estabelecer a prática a partir de certa “tradição histórica” de realização parece um exercício pouco promissor, ainda mais se considerarmos o caráter particular dos métodos de realização dessas obras.

Dessa forma, será realizada a seguir uma análise dos principais trabalhos já realizados. O objetivo dessa análise é buscar perceber traços comuns entre a metodologia utilizada para a sua confecção ou mesmo princípios norteadores que possam ser deduzidos, para que se possa tentar estabelecer características que possam ser consideradas exclusivas do que denominamos “jornalismo online a partir de bases de dados públicas”.

Um dos trabalhos mais importantes feitos a partir de dados públicos é o *site* Everyblock.com<sup>10</sup>, criado pelo americano Adrian Holovaty. Adrian é um dos ainda raros casos de jornalista que também é competente em programação e desenvolvimento *web*, tendo participado da criação do Django, um *framework* para a linguagem *Python*<sup>11</sup>. No jornalismo, atuou em jornais como *The Atlanta Journal-Constitution*, *Lawrence Journal-World* and *The Washington Post*.

---

<sup>10</sup><http://www.everyblock.com/>. Acessado em 8 jul 2011.

<sup>11</sup>De acordo com informações disponíveis em

<https://docs.djangoproject.com/en/dev/internals/committers/#internals-committers> . Acessado em 6 jul 2011.

Em 2005, Holovaty criou o *Chicagocrime.org*, o embrião daquilo que posteriormente se tornaria o projeto *Everyblock*. Realizando um *mashup* das informações disponibilizadas pelo Departamento de Polícia de Chicago com os mapas do *Google Maps*, o *site* permitia que o leitor explorasse os dados sobre os crimes registrados pela autoridade policial de diferentes maneiras (por local, tipo de crime, data, etc.). O projeto ganhou o *Batten Award for Innovations in Journalism* de 2005.

O trabalho se desenvolveu para o *Everyblock*, um agregador de notícias das microrregiões – os quarteirões, ou *blocks* - que compõem cada cidade coberta (16 cidades dos Estados Unidos são cobertas pelo site em julho de 2011 – Seattle, Portland, San Francisco, San Jose, Los Angeles, Dallas, Chicago, Houston, Atlanta, Detroit, Washington D.C., New York City, Boston, Philadelphia, Charlotte e Miami). Notícias e dados relacionados a cada bloco são listados, extraídos de bases de dados públicas ou outras que foram disponibilizadas na *internet* (como lista de imóveis na região), assim como há um espaço para que cidadãos que vivem naquela região ou cidade interajam entre si e os assuntos que estão sendo abordados.

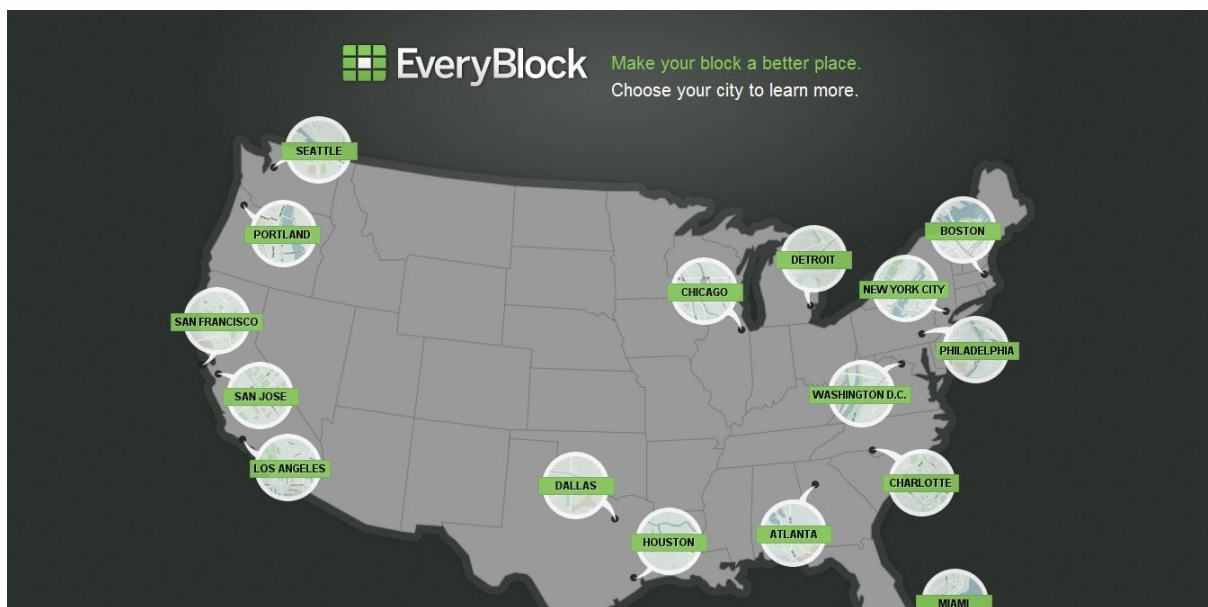


Fig. 1 – Página inicial do *site* Everyblock.com.

O que o *site* *Everyblock* faz é servir como uma interface inteligível para esse montante de dados públicos que são reunidos, com o objetivo de organizar o acesso

do usuário a essas informações de forma lógica – a partir de uma seleção de bairros de uma cidade ou de um determinado tipo de informação que se está procurando. Assim sendo, não é na exclusividade dos dados onde repousa os méritos do *site*, mas na forma na qual esses dados estão organizados e na possibilidade de seleção e intercruzamento desses dados a partir das demandas individuais de cada indivíduo que o acessa. A experiência de informação é específica a cada necessidade e muda a cada acesso por conta disso.

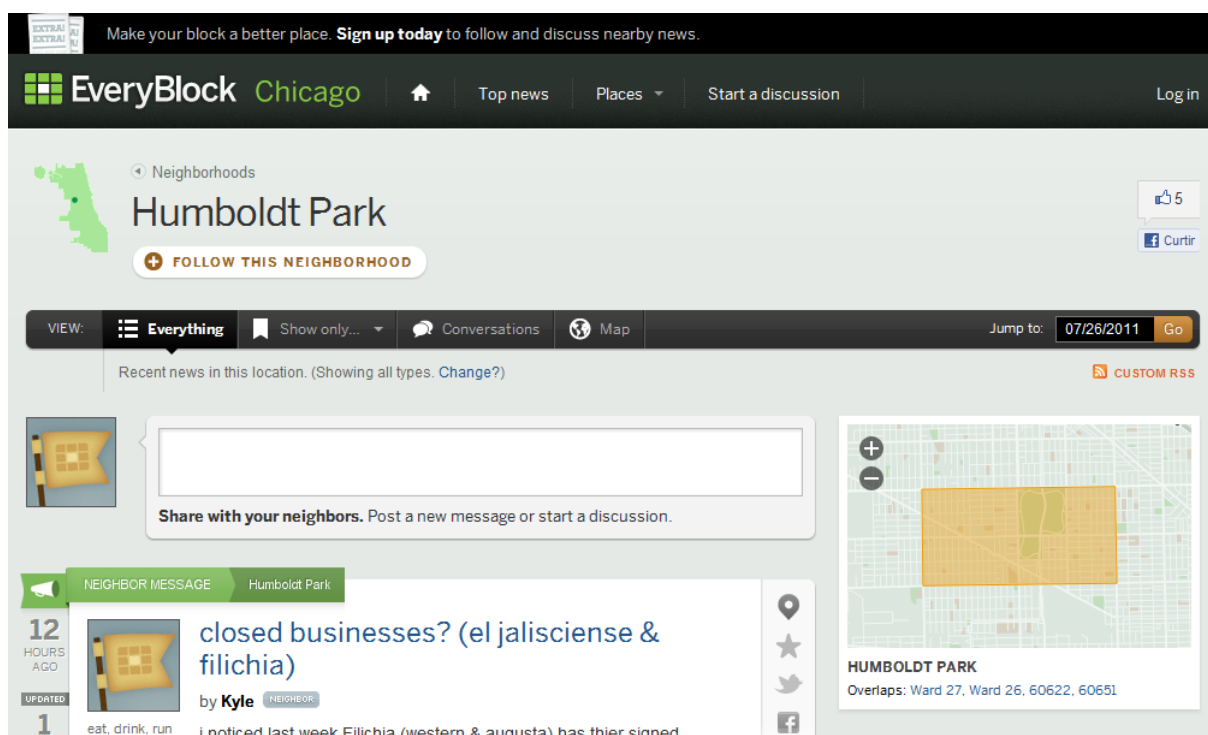


Fig. 2 – Página referente a cidade de Chicago no *site* Everyblock.

HOLOVATY (2006a) acredita que o jornalismo deveria se centrar menos no esquema clássico de coletar as informações e escrever uma história com elas. Ao fazer isso, destilam a informação numa bolha de texto e impede que essas informações sejam utilizadas para outros propósitos – ou seja, sirvam como base para a recombinação e reinterpretação. Segundo ele, não trabalhar a informação como dados é desperdiçar o potencial de utilizar aquela informação da maneira mais interessante para seus leitores, seja na forma de uma história contada através do formato tradicional, com textos e fotos; até recombinações para a construção de mapas interativos ou linhas do tempo, por exemplo.

“Da forma que eu vejo, há três tarefas básicas que os jornalistas fazem:

1 – Coletar informações: Isso envolve falar com fontes, analisar documentos, tirar fotos, etc. Isso é reportagem.

2 – Destilar informações: Isso envolve aplicar julgamento editorial para decidir que partes da informação coletada são importantes e relevantes.

3 – Apresentar informações: Isso envolve modelar a informação destilada em um formato que seja acessível aos leitores. Alguns exemplos: estilo de escrita (pirâmide

invertida, etc.), correção de cor das fotos, design das páginas do jornal.

'Fazer jornalismo através de programação de computadores' é apenas uma maneira diferente de atingir esses objetivos.” (HOLOVATY, 2006a)

O conceito de Holoaty permite que a atividade jornalística seja vista para além das práticas tradicionais e que se leve em consideração aquilo que lhe é essencial e imutável a despeito do suporte que é utilizado: obtenção, organização e apresentação de informações. A programação, nesse caso, é apenas mais uma ferramenta que pode ser utilizada para a realização dessas tarefas e que pode ser dominada pelo jornalista e aplicada quando o meio mais conveniente para que essa mensagem seja compreendida por sua audiência. Na verdade, ele considera que o jornalismo não deveria investir muito tempo neste tipo de questionamento:

“(...) a questão de 'como é esse jornalismo' é acadêmica. Os jornalistas devem se preocupar menos com o que é ou não é 'jornalismo' e mais com informações importantes e focadas que são úteis para a vida das pessoas e as ajudem a compreender o mundo. Um jornal deveria ser isso: um olhar honesto para informações atuais e importantes para os leitores.” (HOLOVATY, *idem*)

Outro exemplo importante do uso de bases de dados públicas na construção de interfaces destinadas à informação do público é o projeto Excelências<sup>12</sup>, concebido pelo jornalista Marcelo Soares e desenvolvida pela ONG Transparência Brasil. O *site* Excelências permite que o usuário pesquise informações sobre os parlamentares em exercício nas casas legislativas estaduais e federais e membros das câmaras municipais das capitais brasileiras.

---

<sup>12</sup>Disponível em <http://www.excelencias.org.br>. Acessado em 31 jul 2011.



Fig. 3 – Página inicial do projeto Excelências (<http://www.excelencias.org.br>)

As informações acessadas são extraídas de bases de dados públicas, disponibilizadas pelos próprios órgãos legislativos ou judiciários (como Tribunal Superior Eleitoral, entre outros). Esses dados podem ser visualizados de diferentes formas: através de uma pesquisa a partir do nome de um parlamentar ou através do cruzamento entre uma das categorias de informação listadas e a casa legislativa desejada. Uma série de informações pode ser acessada a partir desta primeira pesquisa: podem-se ler os projetos de leis apresentados pelos parlamentares a partir de uma pesquisa sobre produção legislativa, por exemplo. Quando necessário, o *site* direciona o usuário para a fonte do conteúdo, normalmente hospedado na página da *internet* oficial do órgão responsável por aquela informação.

As informações vindas de fontes públicas são cruzadas com outras produzidas por outros projetos do Transparência Brasil: o projeto Às Claras gera informações referentes ao financiamento das campanhas eleitorais, permitindo que se visualize a listagem de pessoas físicas e jurídicas que contribuíram para candidatos a cargos da esfera estadual, distrital e federal e a quantia doada. O projeto Deu no Jornal produz uma base de dados de reportagens relacionadas à corrupção, permitindo que seja possível pesquisar as ocorrências de parlamentares citados nessas reportagens.

SOARES (2011) explicou como surgiu a ideia para o projeto Excelências e como ele foi realizado:

“(...) Aí, um dia me chamaram, me ofereceram uma vaga pra ser o coordenador do projeto “Deu no Jornal” do Transparência Brasil, que é um banco de dados de reportagens sobre corrupção publicadas em 50 jornais do Brasil. Tinha uma equipe de três estagiários. Desde o final de 2005, depois que eu deixei de ser gerente da ABRAJI, eu pensei “Bom, agora eu quero fazer jornalismo mesmo, de novo. Quero mexer com banco de dados. Quero fazer alguma coisa que seja transformadora. Mostrar pra que servem dados”.

Então eu comecei a conversar com o Philip Meyer. Eu falo com ele há anos, mas ali eu falei: 'Philip, eu tô afim de fazer alguma coisa com dados. Ano que vem tem eleição. Vai ser uma eleição importante, que o Lula vai concorrer à reeleição, tem os deputados lá e tal. E eu tenho a avaliação que existe uma quantidade inédita de informação pública sobre esses caras. Dá pra fazer um perfil redondo deles. No site da Câmara, tem os dados de presença deles, tem os dados de como é que eles votam nos projetos, tem os dados de como eles gastam verbas indenizatórias. Se pesquisar no Supremo, tem os processos aos quais eles respondem. Se eu pesquisar no TSE, tem os dados de quem bancou a campanha deles. Dá pra fazer um perfil redondo deles'.

(...)

Eu fui trabalhar com o Transparência Brasil, que já tinha alguns bancos de dados, tipo “Às Claras” que é de contribuições eleitorais, e propus ao diretor de lá que a gente fizesse um banco de dados dos deputados. (...) Então a gente bolou um serviço que acabou virando o “Excelências”, que é um banco de dados que existe até hoje, tá muito melhorado, muito ampliado. (...) O 'Excelências' foi lançado em agosto de 2006. (...) A gente começou lançando São Paulo, Rio... Acho que só São Paulo e Rio, eram 150 nomes... Não, foi São Paulo, Rio e Brasília... São Paulo, Rio e Minas. Foram 150 deputados, aí depois cada dia a gente ia adicionando um estado, que eram os novos deputados que entravam e tal. Isso deu uma visibilidade absurda ao Transparência, tanto é que hoje é a espinha dorsal da entidade.”

O “Excelências” fez com que Marcelo Soares e o Transparência Brasil vencessem o 56º Prêmio Esso de Jornalismo na categoria “Melhor Contribuição à Imprensa”, no ano de 2006<sup>13</sup>. A escala do projeto, que lida com grandes números tanto em acessos quanto nas diversas bases de dados que são integradas através

---

<sup>13</sup>Conforme informação disponível em

[http://www.premioesso.com.br/site/premio\\_principal/index.aspx?year=2006](http://www.premioesso.com.br/site/premio_principal/index.aspx?year=2006). Acessado em 31 jul 2011.

da interface de navegação, somado ao pioneirismo da iniciativa faz com que este seja um dos trabalhos mais representativos realizados no Brasil sobre o uso de bases de dados públicas para a construção de um conteúdo jornalístico para mídias digitais.

O uso jornalístico das bases de dados públicas não se limita a estruturação de interfaces de recombinação de informações. O especial “*Faces of the Dead*”, do site do jornal americano *The New York Times* é um exemplo no qual as informações obtidas através das bases de dados públicas servem como fonte e suporte para criar uma narrativa.

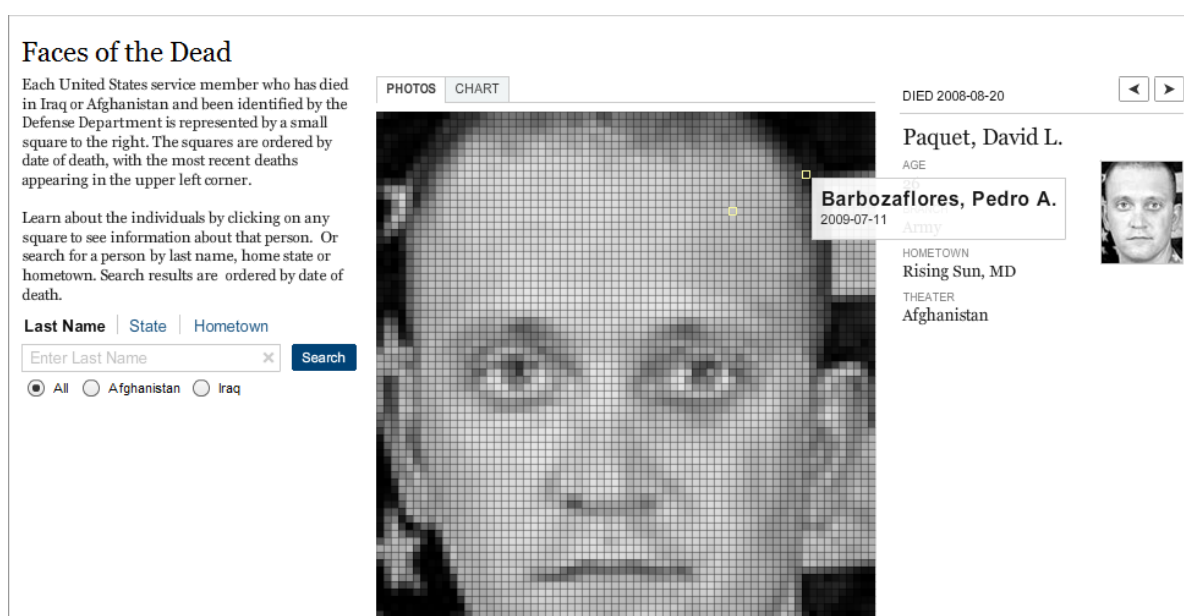


Fig. 4 – Página inicial do especial “*Faces of the Dead*”, do site do jornal *The New York Times*.

Este especial foi criada por uma equipe de repórteres composta por Gabriel Dance, Aron Pilhofer, Andy Lehren e Jeff Damens. Eles construíram um mosaico com fotos dos soldados americanos mortos nas operações do Iraque e do Afeganistão que puderam ser identificados pelo Departamento de Defesa dos Estados Unidos. Cada foto é acompanhada pelo nome do soldado, idade, cidade de nascimento, seção na qual servia (Exército, Marinha ou Aeronáutica) teatro de atuação (Iraque ou Afeganistão) e data do falecimento.

Na coluna à esquerda do mosaico, é possível pesquisar através do sobrenome do soldado, o estado ou a cidade natal. O usuário pode filtrar sua pesquisa pelo teatro de operações, se quiser. É possível escolher uma visualização que mostra, ao invés do mosaico composto pelas fotografias dos soldados, uma espécie de gráfico. Duas cores diferentes identificam o teatro de operações onde ocorreu a morte representada por cada pequeno quadrado, permitindo que se veja a concentração de falecimentos através do segundo mosaico resultante da composição, conforme mostrado na FIG. 5 (a seguir).

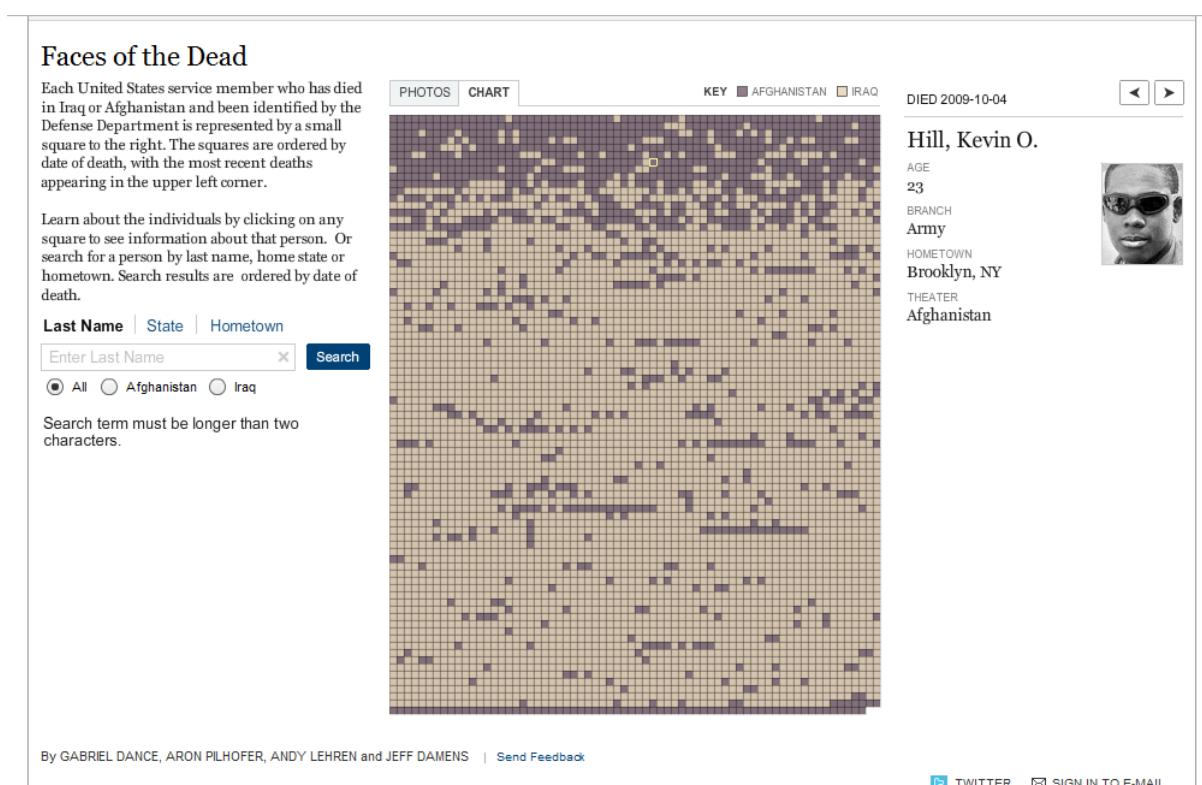


Fig. 5 – Gráfico de mortes por teatro de operações do especial “Faces of the Dead”.

Os dados públicos utilizados servem de uma forma bastante única para a narrativa jornalística em relação a outras reportagens similares sobre militares mortos nestes conflitos. As baixas não são apenas números; a partir das informações disponibilizadas pelo Departamento de Defesa, cada vítima norte-americana resultante dos conflitos ganha um rosto e uma identidade, ao invés de ser massificada como números em um relatório. Ao construir o especial como uma interface para essa base de dados, permite-se que o usuário tenha uma experiência individual, a partir de sua exploração da estrutura, com cada soldado que faleceu.

Isso faz com que a narrativa jornalística tenha um impacto emocional maior, numa tentativa de aproximar o leitor da experiência da perda de cada uma daquelas vidas. Essa experiência só é possível nesse grau de intimidade porque o ambiente digital no qual o especial foi desenvolvido permite que uma base de dados muito grande seja apresentada plenamente aberta à exploração, sem a limitação do espaço que tornaria necessária uma edição e, portanto, uma experiência muito mais limitada se a mesma coisa fosse tentada num meio como o jornal impresso ou a televisão.

Esta seleção demonstrar como a utilização de bases de dados públicos pelo jornalismo online assume diversas formas a partir de trabalhos realizados por jornalistas especializados. É possível notar que não existe apenas um tipo de tema a ser tratado por essa prática. Ela é, na verdade, um modo de contar uma história que serve muito bem ao jornalismo, pois se apoia em informações concretas e permitem a exploração do leitor através das informações que apresenta, ao invés de uma condução passiva pela história. Isso faz com que a técnica permita a construção de narrativas muito ricas, pois permite um grande número de caminhos a serem percorridos pelo leitor.

No capítulo seguinte, trataremos da natureza das bases de dados públicas que servem a esses e outros exemplos de narrativas jornalísticas *online*. O intuito é compreender o que as compõem e porque podem ser importantes para o desenvolvimento do jornalismo nesse ambiente das mídias digitais.

## **2. Bases de dados públicos e informação pública**

Entender no que consistem o que se convencionou chamar de bases de dados públicos é essencial para poder tentar compreender as relações existentes para com o jornalismo, com ênfase no produzido para os meios digitais. Neste capítulo, pretende-se estabelecer as definições do que são bases de dados públicas e quais são os tipos de informações que as compõem. Em seguida, pretende-se estabelecer o conceito de transparência governamental. Serão abordados quais são as características propostas para que os dados governamentais sejam considerados abertos e sua importância para o jornalismo produzido para mídias digitais.

### **2.1 - Bases de dados públicas e transparência**

Transparência, no campo da política, deve ser entendida como uma propriedade "atribuída a instituições que permitem que os cidadãos 'vejam através' de suas estruturas, e possam assim tomar consciência dos fluxos informacionais que transpassam os centros de poder" (SILVA, 2010, p. 7). De acordo com a pesquisadora, a noção de transparência como visibilidade dos processos de governo pode ser localizada historicamente nos ideais de participação geral do indivíduo na vida pública da Revolução Francesa e na ágora da Grécia Antiga, um espaço público comum de debate que inspirou esse conceito nos revolucionários franceses. De fato, o conceito de transparência e democracia tem a origem muito próxima, conforme demonstra BOBBIO (*apud* SILVA, *idem*): "Com um aparente jogo de palavras, pode-se definir o governo da democracia como o governo do poder público em público" (p. 36). Com isso, pode-se concluir que o autor considera a capacidade de divulgar informações sobre as atividades governamentais de forma ampla e clara como característica definitiva e fundamental de um regime democrático:

"O que distingue o poder democrático do poder autocrático é que apenas o primeiro, por meio da livre crítica e da liceidade de expressão dos diversos pontos de vista, pode desenvolver em si mesmo os anticorpos e consentir formas de desocultamento" (BOBBIO, *apud* VALENTE, 2004, p. 73)

A questão da importância do acesso público a informações governamentais e, por lógica, de pertencimento do público pode ser traçada como uma preocupação de longa data em alguns importantes estados democráticos do mundo. Na Suécia, o acesso a informações públicas é garantido desde 1776, quando o primeiro Ato de Liberdade de Imprensa foi promulgado pelo então Rei Adolfo Frederico<sup>14</sup>. O atual texto desta legislação, datado de 1949, garante o direito dos cidadãos a terem acesso e visualização de documentos e atividades públicas<sup>15</sup>, sendo um dos princípios básicos da legislação daquele país. Para que o documento esteja acessível, basta que seja considerado oficial e que não seja considerado secreto<sup>16</sup>.

Nos EUA, a administração de Barack Obama afirmou a importância do acesso aos dados públicos e sua utilização pela população através de um memorando chamado “*Transparency and Open Government*” (“Transparência e Governo Aberto”), publicado em 21 de janeiro de 2009. Nele, o presidente compromete-se a “criar um nível de abertura no governo sem precedentes” e afirma que “a abertura irá fortalecer nossa democracia e promover eficiência e efetividade no governo”<sup>17</sup>. A partir do memorando, foi lançada a política *Open Government Initiative* (Iniciativa de Governo Aberto) e, posteriormente, o projeto *Data.gov*, um repositório de dados governamentais federais estadunidenses. Essas iniciativas complementam o Freedom of Information Act<sup>18</sup> (FOIA) (Ato de Liberdade de Informação), promulgado em 1966, que postula que qualquer pessoa tem o direito de acesso a registros de agências federais que não sejam protegidos de revelação pública pelas exceções definidas por lei.

O Brasil também possui iniciativas no sentido de tentar tornar os processos

---

<sup>14</sup> A tradução para o inglês do texto original “His Majesty’s Gracious Ordinance Relating to Freedom of Writing and of the Press (1766) pode ser encontrada em ANDERS CHYDENIUS FOUNDATION. The World’s First Freedom of Information Act. Kokkola (Finlândia): Art-Print, 2006. Disponível em <http://www.scribd.com/doc/5885744/The-Worlds-First-Freedom-of-Information-Act-SwedenFinland-1766>. Acessado em 24 mar 2011.

<sup>15</sup> De acordo com informações do artigo “Consultations of the Task Force with other Jurisdictions on their Access to Information Regimes”. Disponível em <http://www.atirtf-geai.gc.ca/consultation-sweden-e.html>. Acessado em 22 mar 2011.

<sup>16</sup> De acordo com o documento “Public Access to Information and Secrecy Act”, publicado pelo Ministério da Justiça da Suécia. Disponível em <http://www.sweden.gov.se/sb/d/11929/a/131397>. Acessado em 22 mar 2010.

<sup>17</sup> De acordo com o memorando “Transparency and Open Government”. Disponível em [http://www.whitehouse.gov/the\\_press\\_office/TransparencyandOpenGovernment/](http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment/). Acessado em 23 mar 2011.

<sup>18</sup> O texto integral está disponível em <http://www.justice.gov/oip/foia-guidance-resources.html>. Acessado em 23 mar 2011.

administrativos governamentais mais claros para os cidadãos. O país possui alguns dispositivos legais destinados a definir os parâmetros para a divulgação de dados referentes às atividades dos diversos órgãos públicos. A principal regulamentação nesse sentido se encontra justamente no artigo 5º, inciso XXXIII e pelo artigo 37 da Constituição, conforme reproduzidos a seguir:

(...) Art. 5º, Inciso XXXIII - **todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral**, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado;

(...) Art. 37 - A administração pública direta e indireta de qualquer dos Poderes da União, dos Estados, do Distrito Federal e dos Municípios obedecerá aos princípios de legalidade, impessoalidade, moralidade, **publicidade** e eficiência e, também, ao seguinte: (...) [grifos deste pesquisador]

Em meados de 2005, iniciaram-se algumas movimentações de organizações civis pela criação de uma regulamentação mais clara sobre a divulgação das informações administrativas, com clara ênfase aos dados orçamentários, para orientar de que forma deveria se dar essa publicidade dos atos da administração pública, conforme previsto pela Constituição.<sup>19</sup> Em maio de 2009, o presidente da República Luiz Inácio da Silva sancionou a Lei Complementar nº 131, que acrescenta dispositivos à Lei Complementar nº 101 que legisla sobre a publicidade das finanças públicas, entre outras coisas. A Lei Complementar nº 131<sup>20</sup> define os parâmetros de divulgação das informações, conforme se segue:

“Art. 1º O art. 48 da Lei Complementar nº 101, de 4 de maio de 2000, passa a vigorar com a seguinte redação: (...)

Parágrafo único. A transparência será assegurada também mediante:

(...) II – liberação ao pleno conhecimento e

<sup>19</sup> Conforme descrito por ABRAMO, Cláudio Weber. *A César o que é de César (mas devagar com o andor)*. Artigo publicado no site Observatório da Imprensa em 7 abr 2009. Disponível em <http://www.observatoriodaimprensa.com.br/artigos.asp?cod=532CID003>. Acessado em 23 mar 2011.

<sup>20</sup> O texto integral da Lei Complementar nº 131, de 27 de maio de 2009, do qual as citações foram tiradas, está disponível em [http://www.planalto.gov.br/ccivil\\_03/Leis/LCP/Lcp131.htm](http://www.planalto.gov.br/ccivil_03/Leis/LCP/Lcp131.htm). Acessado em 24 mar 2011.

acompanhamento da sociedade, em tempo real, de informações pormenorizadas sobre a execução orçamentária e financeira, em meios eletrônicos de acesso público; (...)"

A legislação também define que tipo de dados deverão ser disponibilizados:

"Art. 2º A Lei Complementar nº 101, de 4 de maio de 2000, passa a vigorar acrescida dos seguintes arts. 48-A, 73-A, 73-B e 73-C:

"Art. 48-A. Para os fins a que se refere o inciso II do parágrafo único do art. 48, os entes da Federação disponibilizarão a qualquer pessoa física ou jurídica o acesso a informações referentes a:

I – quanto à despesa: todos os atos praticados pelas unidades gestoras no decorrer da execução da despesa, no momento de sua realização, com a disponibilização mínima dos dados referentes ao número do correspondente processo, ao bem fornecido ou ao serviço prestado, à pessoa física ou jurídica beneficiária do pagamento e, quando for o caso, ao procedimento licitatório realizado;

II – quanto à receita: o lançamento e o recebimento de toda a receita das unidades gestoras, inclusive referente a recursos extraordinários."

É importante notar que a expressão "tempo real" que é usada para a definição sobre a disponibilidade pela qual as informações deverão estar disponíveis ao público. A definição do dicionário o define como "o tipo de processamento de resultado imediato, quase simultâneo às ações que o acionam"<sup>21</sup>. Isso estaria em consonância com as possibilidades apresentadas pelas TIC's para o compartilhamento instantâneo de informação. No entanto, a interpretação descrita por FILGUEIRAS (2010), o "tempo real" seria juridicamente regulamentado pelo Decreto Federal nº 7.185<sup>22</sup>, que define:

"Art. 1º A transparência da gestão fiscal dos entes da Federação referidos no art. 1º, § 3º, da Lei Complementar nº 101, de 4 de maio de 2000, será assegurada mediante a observância do disposto no art. 48, parágrafo único, da referida Lei e das normas estabelecidas neste Decreto.

---

<sup>21</sup> De acordo com definição do DICIONÁRIO CALDAS AULETE DIGITAL.

<sup>22</sup> O texto integral do Decreto Federal nº 7.185, de 27 de maio de 2010, do qual as citações foram tiradas, está disponível em [http://www.planalto.gov.br/ccivil\\_03/\\_Ato2007-2010/2010/Decreto/D7185.htm](http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2010/Decreto/D7185.htm). Acessado em 24 mar 2011.

Art. 2º O sistema integrado de administração financeira e controle utilizado no âmbito de cada ente da Federação, doravante denominado SISTEMA, deverá permitir a liberação em tempo real das informações pormenorizadas sobre a execução orçamentária e financeira das unidades gestoras, referentes à receita e à despesa, com a abertura mínima estabelecida neste Decreto, bem como o registro contábil tempestivo dos atos e fatos que afetam ou possam afetar o patrimônio da entidade.

§ 2º Para fins deste Decreto, entende-se por:

II - liberação em tempo real: a disponibilização das informações, em meio eletrônico que possibilite amplo acesso público, até o primeiro dia útil subsequente à data do registro contábil no respectivo SISTEMA, sem prejuízo do desempenho e da preservação das rotinas de segurança operacional necessários ao seu pleno funcionamento; (...)"

Visando adequar-se às leis, as três esferas do poder brasileiro – municipal, estadual e federal – lançaram portais onde disponibilizam os dados relacionados às finanças públicas. As informações relativas ao Poder Executivo Federal estão disponíveis através da internet no Portal da Transparência do Governo Federal<sup>23</sup>. De forma similar, órgãos estaduais e municipais como a Secretaria de Fazenda do Estado de São Paulo<sup>24</sup> e a Prefeitura de São Paulo<sup>25</sup> também possuem *websites* ou seções em seus *websites* destinados a disponibilizar publicamente relatórios de despesas e arrecadações.

É claro que a divulgação de dados relativos ao orçamento é apenas uma pequena parte do grande montante de informação possuída pelos órgãos governamentais. De acordo com SILVA (2011), historicamente se compreende que, ao falar de disponibilização de informações públicas para a sociedade, entende-se que se está tratando de informações referentes ao uso do dinheiro público. De acordo com a compreensão da jornalista, esses dados não são tão importantes quanto outros que também são produzidos pela administração pública e ainda não são completamente disponibilizados.

---

<sup>23</sup> Disponível em <http://www.portaltransparencia.gov.br/>. Acessado em 21 out 2010.

<sup>24</sup> Disponível em <http://www.fazenda.sp.gov.br/contas1/>. Acessado em 21 out 2010.

<sup>25</sup> Disponível em <http://deolhonascontas.prefeitura.sp.gov.br/>. Acessado em 21 out 2010.

“(...) Eu me preocupo menos com ter dados disponíveis para fazer controle social e mais [com ter] dados disponíveis para fazer participação política. Então, me interessa bastante, por exemplo, saber dados da saúde, dados da educação, dados de serviço público; porque dessa forma eu posso não apenas fiscalizar como esse serviço funciona, mas atuar de um jeito mais estratégico enquanto cidadão, por exemplo.” (SILVA, *idem*, p. 11)

Em novembro de 2011, foi sancionada a lei nº 12.527/2011, conhecida como lei de acesso a informações públicas. O projeto apresentado em abril de 2010 na Câmara dos Deputados e aprovado em regime de urgência. Antes que isso acontecesse, houveram tentativas para que o texto do projeto fosse alterado substancialmente, de modo a permitir o sigilo por tempo indefinido de determinados documentos públicos, fato que desfiguraria por completo uma legislação destinada justamente a permitir o acesso públicos às informações. Após uma extensa negociação e pressão de diversos setores da sociedade, o Senado aprovou o texto sem alterações e encaminhou para a sanção da Presidência da República em outubro de 2011.

O projeto regula o acesso a informações previsto pelos artigos 5º, 37 e 216 da Constituição da República Federativa do Brasil. Declara que cabe aos órgãos e entidades públicas assegurar o acesso às informações relativas a diversos atividades da administração pública como o a implementação e resultado de programas e ações e auditorias e prestações de contas, entre outras. Outro ponto importante do projeto de lei é que as informações devem ser disponibilizadas de duas maneiras: deve informar, rotineiramente e de forma automática, suas competências, organograma e contatos, sua movimentação financeira e despesas, informações sobre procedimentos licitatórios, dados gerais para acompanhamento de políticas e obras públicas e respostas a perguntas frequentes da sociedade. Além disso, todo órgão público deve manter uma página eletrônica na internet onde, anualmente, disponibilizará a lista de documentos que deixaram de ser sigilosos, a lista de documentos que são sigilosos e um relatório estatístico sobre pedidos de informação recebidos, atendidos e indeferidos.

A situação do acesso a informações em outras esferas do poder público já se encontra mais desenvolvida. A publicidade dos chamados atos processuais na

Justiça brasileira é uma medida também prevista pela Constituição Federal e no Código de Processo Civil brasileiro, com a exceção dos casos em que correm excepcionalmente sob sigilo de justiça. Ou seja, é garantido que o processo possa ser acessado e conhecido tanto pelas partes quanto pelo público, excetuando-se casos em que esse acesso ao processo possa causar algum dano ou constrangimento a alguma das partes, como numa separação litigiosa, por exemplo.

## 2.2 Acesso e possibilidade de uso dos dados públicos

SILVA (2010) acredita que “a maneira como os governos disponibilizam suas informações deve permitir que a inteligência coletiva crie melhores formas de trabalhá-las do que os próprios governos poderiam fazer” (p. 121). A autora acredita que tais informações, para cumprir esta função social, devem se adequar ao conceito de “dados abertos”. A definição adotada por ela foi cunhada a partir de um encontro entre 30 ativistas pela “governança aberta” durante um encontro em Sebastopol, Califórnia.

De acordo com a proposta desse grupo, os dados governamentais podem ser considerados “abertos” quando apresentam as seguintes características:

<b>Completos</b>	Todos os dados públicos são disponibilizados. Dados públicos são dados que não se submetem a limitações válidas de privacidade, de segurança ou de privilégio.
<b>Primários</b>	Os dados são coletados na sua fonte, com o maior nível possível de granularidade, não estando em formas agregadas ou modificadas.
<b>Atualizados</b>	Os dados são disponibilizados tão rápido quanto seja necessário para preservar seu valor.
<b>Acessíveis</b>	Os dados estão disponíveis para o maior escopo possível de usuários e para o maior escopo possível de finalidades.
<b>Legíveis por máquina</b>	Os dados estão razoavelmente estruturados para permitir processamento automático.
<b>Não-discriminatórios</b>	Os dados estão disponíveis para todos, sem necessidade de registro para acessá-los.
<b>Não-proprietários</b>	Os dados são disponibilizados num formato sobre o qual nenhuma entidade tem controle exclusivo.
<b>Livres de licenças</b>	Os dados não estão sujeitos a nenhuma forma de direito autoral, patente, propriedade intelectual ou sigilo industrial. Restrições razoáveis de privacidade, de segurança e de privilégio podem ser permitidas.

Tabela 1: Princípios dos dados governamentais abertos, conforme SILVA (*ibidem*, p. 117)

Esses princípios foram definidos com o objetivo de que tais dados estejam abertos para o maior número possível de pessoas e para o maior número possível de usos. Como se tratam de informações governamentais, considera-se que elas sejam de pertencimento de todos os indivíduos daquela determinada nação e que não podem estar sujeito a limitações, excetuando aquelas relacionadas à informações particulares ou que possam ameaçar de alguma forma a soberania ou a segurança da nação.

É interessante notar que a origem das ações com o objetivo de ampliar a disponibilidade de dados públicos através dos meios digitais é, na verdade, extensões dos objetivos e ações de movimentos bem mais antigos, organizados no mundo inteiro com o objetivo de tornar os processos administrativos públicos mais claros para as suas respectivas sociedades. Estes grupos perceberam o potencial da *internet* de servir como o modo mais fácil, rápido e inteligente de permitir que qualquer cidadão que deseje o acesso a esse tipo de dado possa obtê-lo.

No cruzamento da luta por um governo mais transparente e na ampliação das utilizações políticas da *internet* pelo cidadão surgem iniciativas como a Esfera e os eventos “*Transparência Hack Day*”. A Esfera define-se como:

“(…) um projeto (em fase bem inicial) de Daniela Silva e Pedro Markun, com foco em comunicação, política e novas tecnologias. Queremos ver acontecer, na política, a mesma revolução de práticas, valores e princípios que as novas tecnologias trouxeram para a comunicação. Para isso, agimos em prol da transparência, da abertura e da colaboração, por meio de projetos focados no empoderamento do cidadão, de ações de formação de cultura ligadas a esses valores e outras iniciativas. Nossa intenção é fomentar a ideia de que um **governo aberto** pode ser potencializado pela dinâmica das redes.”<sup>26</sup>[grifo deste pesquisador]

Se os movimentos por governos abertos que conduziram seus esforços para a abertura de dados através da *web* forem tomados como uma espécie de “1ª geração”, o projeto Esfera pode ser considerado como uma “2ª geração”, que se desenvolve num fluxo contrário ao primeiro, se originando a partir das experiências

---

<sup>26</sup>Informações obtidas no *site* do projeto Esfera. Disponível em <http://blog.esfera.mobi/about/> . Acessado em 31 jul 2011.

com a *internet* e as demais tecnologias de comunicação digitais e que percebem a necessidade de utilização desses meios como potencializadores da ação política do cidadão através do seu uso para o acesso a informações públicas. Daí o caráter horizontal da organização e dos eventos promovidos por eles, dos quais o “*Transparência Hack Day*” pode ser considerado o mais representativo. A primeira realização está intimamente ligada com o próprio surgimento do projeto Esfera, conforme explicado por SILVA (2011):

“(...) Em janeiro de 2009, eu e o Pedro Markun – que é meu sócio – (...) começou a desenhar o que ia ser a Esfera muito 'periféricamente' ainda, então listando princípios: 'a gente quer trabalhar com isso, não quer trabalhar com isso'. A gente sabia que era política, *internet*, mas política no entendimento do empoderamento do cidadão, então política do lado da sociedade. E aí, em outubro, a primeira ação da Esfera aconteceu, que foi o *Transparência Hack Day*.

Antes disso ainda teve o clone do *blog* do Planalto, que foi na verdade o estopim de tudo. Foi uma brincadeira, foi uma ideia de fazer uma provocação para a *internet*, que tava reclamando e chorando porque o tal *blog* do Planalto não tinha comentário, mas tinha um monte de possibilidades de interação e a gente estava querendo mostrar que 'olha, de repente a gente tem possibilidades de ação aqui que não são só verbais, que não é só falar, que não é só expressar o nosso descontentamento. (...)

Considerando que essa mobilização [gerada pelo clone do *blog* do Planalto] estava bacana, veio o primeiro *Transparência Hack Day*. (...) que veio logo na sequência. Foi logo, tipo, um mês ou quase dois meses depois da história do clone. Que foi essa proposta de encontro. (...) Trazer jornalistas e desenvolvedores, eram esses dois grupos que estavam na mente no começo. Mas aí, de jornalista virou acadêmico, ativista, gente de governo, gente de ONG, cidadãos interessados. E veio desenvolvedores, gente que é nativo digital, gente da Rede pra trabalhar junto com dado público na Rede. Então, o primeiro *Transparência Hack Day* foi um convite pra juntar desenvolvedores, pessoas que normalmente fazem *Hack Days*, que é sentar e usar a vontade de escrever códigos para cumprir desafios; (...)

(...) [Fizemos] o convite na Rede e assim foi feito. Apareceram cem pessoas aqui, dessas cem pessoas começou a se formar uma comunidade que hoje tem mais de quatrocentos membros e que continua fazendo encontros como esse de um jeito meio anárquico. (...) Então, tem *Transparência Hack Day* que acontece em São Paulo e, ao mesmo tempo, tem em BH, tem em Brasília, tem no DF, tem no Rio, tem no Pará. (...) A gente ainda tem que, acho, “espalhar o vírus” (...)

Justamente por causa dessa natureza altamente descentralizada que caracteriza a realização desses encontros, não é possível estabelecer quantos já foram realizados. O movimento parece ganhar mais força à medida que atrai mais interessados tanto em seus encontros e grupos de discussão para o debate sobre as ideias quanto também atrai o olhar da mídia para suas ações. O *thackers* já foram pauta de várias matérias em grandes portais de notícia, como o iG e o *site* jornal O Estado de S. Paulo<sup>27</sup>.

Embora conte com a participação de jornalistas, é interessante notar que todo o movimento se caracteriza pela busca do acesso às informações governamentais sem intermediários, tanto governamentais quanto midiáticos. A intenção é permitir que qualquer cidadão possa interpretar e reutilizar esses dados com o objetivo de melhorar a própria sociedade a qual ele pertence.

Ainda assim, a apropriação desses dados não é automática, mesmo que todos os dados sejam disponibilizados de forma irrestrita. É preciso considerar que, para explorar o pleno potencial de todo o montante de informação governamental que existe, é primeiro superar a barreira da técnica. É preciso manipular essa informação, organizá-la e relacioná-la. É certo que o desenvolvimento tecnológico e o trabalho dos diversos grupos de inclusão digital e de democratização da informação permitem que um número cada vez maior de indivíduos, no mundo todo, seja capaz de manipular tais dados brutos e se apropriarem deles de modo a fazer uso, seja de forma particular, ou seja, através do desenvolvimento de aplicações para o uso geral.

Deve-se considerar, todavia, que tal uso exige um investimento considerável de tempo na manipulação direta das informações para a produção de novos sentidos e na aprendizagem necessária para que se seja capaz de fazer isso, sem contar a exigência de acesso ao conhecimento especializado. O jornalismo pode ser

---

<sup>27</sup>A título de exemplo, algumas dessas matérias podem ser encontradas em <http://www.tiagodoria.ig.com.br/2009/10/05/o-que-aconteceu-no-1%C2%BA-transparencia-hackday/> e <http://blogs.estadao.com.br/link/tag/transparencia-hack-day/>. A busca pelo termo “Transparência Hack Day” e “ônibus hacker”, outra ação do grupo, traz dezenas de referências em *sites* de jornais e revistas brasileiros, que não serão relacionados com o intuito de preservar a clareza do trabalho.

um elemento importante na equação formada pela informação pública e o público, atuando como intermediário e auxiliando na compreensão dessa enorme quantidade de informação, embora não seja mais o único meio de acesso a elas.

### **3. PRÁTICAS E RUMOS DO USO DE BASES DE DADOS NO JORNALISMO**

Nos capítulos anteriores, estabeleceu-se a forma com a qual o que denominamos Jornalismo a partir de Base de Dados foi desenvolvido, a partir das experiências anteriores com o Jornalismo de Precisão e a Reportagem Assistida por Computador (RAC). Em seguida, tratamos da questão das bases de dados públicas. Analisamos de que forma as informações governamentais são disponibilizadas para a sociedade tanto no Brasil quanto no resto do mundo. Examinamos o conceito de transparência governamental, sua importância para o acesso às informações relacionadas à administração pública e o desenvolvimento de movimentos em prol desse tipo de abertura.

Neste capítulo, estes estudos serão reunidos com o objetivo de compor o atual cenário das práticas relacionadas ao uso dessas bases de dados públicas para a produção do jornalismo *online*. A partir da observação do que já foi produzido, pretende-se definir algumas das principais características dessa prática e do papel dos dados públicos para sua realização e desenvolvimento. Finalmente, será avaliada a contribuição desse tipo de prática para o desenvolvimento do jornalismo produzido para os meios digitais.

#### **3.1 A estrutura das reportagens digitais baseadas em dados públicos**

O uso de dados não é uma novidade para o jornalismo. Pelo contrário, pode ser considerada uma de suas mais importantes práticas. Ao apurar informações sobre fatos ou realizar entrevistas, o jornalista está justamente manipulando e organizando bases de dados relacionados ao tema com o qual está trabalhando. Então, o que há de peculiar ou específico nos trabalhos estudados nesta pesquisa que os diferem do uso habitual dos dados por parte do jornalismo?

Em primeiro lugar, os trabalhos jornalísticos que foram tomados como referências para essa pesquisa lidam diretamente com dados primários e, a partir deles, produzem informações jornalísticas. Apesar de lidar com dados desde sempre, a prática mais tradicional do jornalismo normalmente se detém em dados

secundários fornecidos pelas organizações (governos, entidades de pesquisa, etc.) e, a partir deles é que constroem a narrativa. Isso significa que esses dados já foram previamente interpretados antes de chegarem às mãos do jornalista e, dessa forma, os resultados do trabalho dependeram do tipo de interpretação que foi dada por aquele que forneceu aquele material.

Ao utilizar dados públicos abertos e primários, o jornalista tem uma liberdade muito maior para produzir. Como está lidando com as informações em estado bruto, a informação resultante corre menos riscos de terem sido distorcidas por interesses de terceiros. Além disso, as informações resultantes da manipulação e filtragem desses dados dependerão basicamente da habilidade e recursos do jornalista em manipulá-los e cruzá-los com outros dados para obter novas informações.

É justamente na forma de manipular e apresentar esses dados onde o uso de bases de dados públicas pelo jornalismo *online* apresenta suas mais interessantes contribuições para o desenvolvimento do jornalismo em meios digitais e, em última instância, para o público. LORENZ (2010) vê o jornalismo a partir de base de dados como “um fluxo de trabalho” (p. 10). Esse fluxo de trabalho permite que os dados sirvam como base para análise visualização e narrativa, o que o autor considera como mais importante nisso tudo. O diagrama (Fig. 6) demonstra como a forma como o jornalismo, sobretudo o produzido para meios digitais, altera sua relação com as fontes de informação. Com isso, também deve alterar a forma de lidar com essas informações, enfatizando seu papel como interpretador dos dados primários.

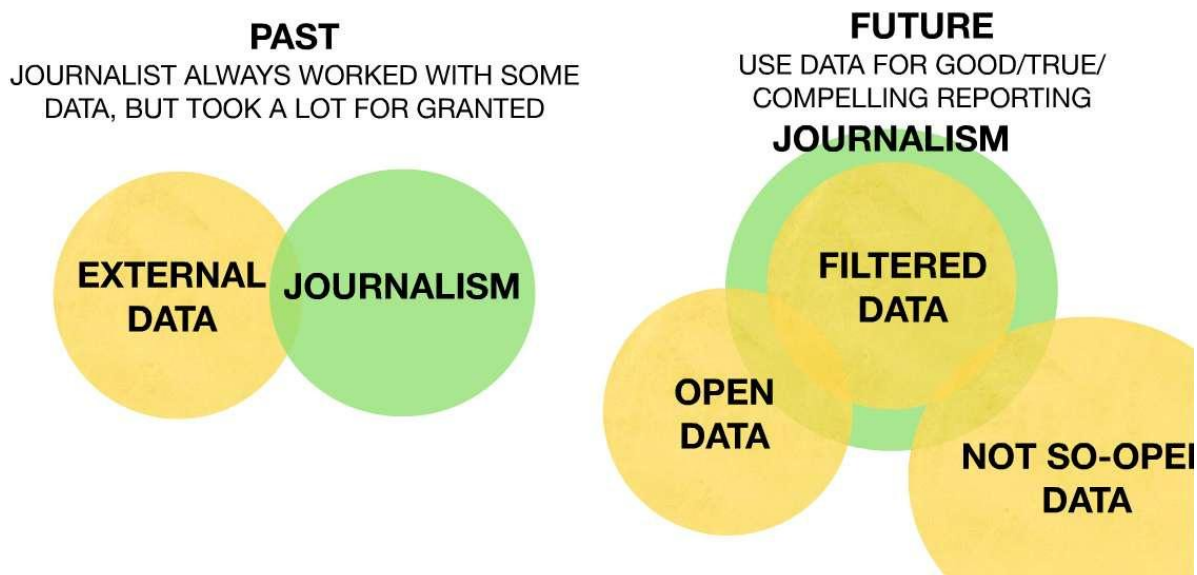


Fig. 6 – Alteração na relação entre jornalismo e dados primários por LORENZ (2010).

Ao utilizar as ferramentas providas pela computação e pelo ambiente de comunicação em rede aliados ao treinamento correto, um jornalista pode trabalhar e processar um volume cada vez maior de dados. Estes dados, por sua vez, estão sendo disponibilizados num volume crescente e de forma mais fácil, graças a movimentos e pressões realizadas pela própria sociedade que exigem transparência nos processos públicos. SOARES (2011) exemplifica como esse acesso mais fácil às informações públicas e o desenvolvimento das tecnologias de comunicação permitiram a execução de projetos que antes seriam impossíveis:

“(...) Vinte e poucos anos atrás, quando a Constituição saiu, um amigo meu, Sérgio Gomes, (...) editou um material chamado “Quem foi quem na constituinte”. Isso foi imensamente importante na eleição de 1989. Esse material foi editado pelo DIAP, Departamento Intersindical de Análises Parlamentares. E eles analisavam como cada constituinte votou nas questões de interesse da classe trabalhadora: salário-mínimo, seguro-desemprego, FGTS e tal. Eles faziam um *ranking* dos deputados com base nisso aí. Isso são dados. Na época, eles precisaram mandar um cara, pagar um cara, pra mandar pra Brasília, para ficar dias e dias, pegando os dados de cada votação, de cada deputado pra marcar se ele votou “sim” ou “não” em cada projeto. Isso tinha um custo absurdo, não era qualquer um que poderia fazer. O DIAP tinha recursos pra isso. (...)

Em 2002, o Fernando Rodrigues, da Folha de S. Paulo, teve uma ideia genial de bater na porta de todos os TREs do Brasil e pedir todas as declarações de bens

de todos os políticos eleitos. Isso demandava custos, precisava contratar um repórter em cada capital (...), precisava bater na porta do TRE, convencer o TRE a fornecer esses dados, alguns não forneceram; precisava pagar as cópias. [A equipe] pegava o documento físico na mão e tirava *xerox*, mandava pro Fernando e ele digitalizava. Ele fez o projeto Controle Público, na época, o que hoje é o 'Políticos do Brasil'. Isso era uma coisa que só uma organização do tamanho da Folha poderia fazer.

Não era possível, não era fácil. Em primeiro lugar, porque precisava convencer a Justiça Eleitoral a fornecer os dados. (...) Pagar esse pessoal nas capitais, pagar as cópias, pagar o correio, pagar o cara que ia digitalizar, que ia digitar isso aí. É um custo absurdo, que muito poucas organizações poderiam arcar. (...) Em 2006, quatro anos depois, na Transparência Brasil, sem um centavo extra além das doações, além do que o Transparência Brasil já tinha pra trabalhar (...) a gente conseguiu botar [o projeto Excelências] no ar, só com informação que já estava disponível.”

Outros pesquisadores também encontraram indícios que há uma relação próxima entre a disponibilidade do acesso aos dados públicos e o desenvolvimento dessa técnica em particular. CAIRO (2008) estudou o que ele define como *visualização analítica* na imprensa, a utilização de informações visuais com o objetivo de servir como apoio a compreensão, aumentando a capacidade de interpretação dos fatos por parte do leitor. Ele acredita que o crescimento do uso desse tipo de visualização por parte da imprensa nos EUA pode ser explicado por alguns fatores:

“o crescente acesso a **bases de dados públicas** de agências governamentais e organizações privadas, a consolidação dos valores do **jornalismo de precisão** e a incorporação de **novos perfis profissionais** (*computer-assisted reporters*, programadores multimídia, cartógrafos, etc.) às redações” (CAIRO, *idem*, p.57-58)

Com a disponibilidade crescente de dados e a chegada de profissionais cujos perfis não se limitam às escolas de jornalismo, mas sim que incorporam o domínio de habilidades relacionados com a operação e programação de computadores, há um desenvolvimento de uma nova forma de produzir informação jornalística. É razoável afirmar que um processo similar acontece no desenvolvimento desta prática específica de jornalismo a partir de bases de dados públicas, como atestam o currículo de diversos profissionais importantes nessa prática como Holovaty e

Marcelo Soares.

Esse perfil se mostra necessário porque as há outras exigências para a construção de uma reportagem deste molde, além das usualmente relacionadas como ser capaz de entrevistar e produzir conteúdos escritos e audiovisuais. Ele deve ser capaz de localizar e obter os dados corretos e relevantes para sua reportagem; ele deve ser capaz de tratar esses dados obtidos de forma a extrair a informação realmente relevante para seu trabalho: “Os dados só se transformam em informação se corretamente filtrados, analisados e contextualizados.” (ABRAJI, p.7). Tomando de empréstimo um termo comumente utilizado em ciências da computação, ele deve ser capaz de estabelecer o *algoritmo* (MANOVICH, *ibidem*, p. 223) correto para que o relacionamento entre os dados obtidos produza um novo resultado, uma nova informação antes indisponível. Como algoritmo, entende-se como um relacionamento lógico entre os dados e que resulta num novo dado após seu processamento.

Tome-se como exemplo um trabalho jornalístico pode pegar bases de dados públicas de mapeamentos de crimes em determinada cidade e de apreensões de drogas na mesma região. Embora relevantes como conjuntos de informações separadas, se relacionadas através de uma sobreposição, essas bases de dados poderiam revelar que áreas de grandes apreensões de drogas também apresentam altos índices de homicídios. Os dados que compõem essas bases foram relacionados de forma lógica – no caso, sobrepostos em busca de coincidências. Na verdade, o próprio ato de selecionar determinadas bases como mais adequadas para a geração de um determinado resultado com o objetivo de produzir novas informações já pode ser considerado um exemplo desta lógica do estabelecimento de um “algoritmo informacional”.

O produtor deste tipo de trabalho também deve fazer com que seu leitor consiga acessar e compreender essa teia de relações informacionais que ele foi capaz de estabelecer. Ele faz isso através da construção de uma *interface*. A interface será o meio que permitirá o acesso e compreensão do usuário sobre a natureza dos dados utilizados e do algoritmo informacional estabelecido. A natureza

dessa interface pode ou não permitir que o leitor interaja com a forma de interação ou de relação entre os dados que a compõem, pode se servir de multimeios para sua composição: tais decisões devem levar em conta a melhor maneira de permitir que os dados obtidos e o algoritmo encontrado possam ser inteligíveis para o leitor, seu consumidor final.

Finalmente, tais estruturas de composição de uma reportagem característica do jornalismo *online* produzido a partir de base de dados públicas são envolvidas por uma última camada, a narrativa. É a narrativa que tem a função de servir como uma espécie de “âncora” significativa que relaciona as informações estruturadas a partir dos dados com pessoas, grupos ou situações que efetivamente representam algo para ele. A produção de uma narrativa ameniza o aspecto asséptico da informação construída a partir das bases de dados, aumentando a proximidade e serve como uma espécie de metáfora, dando rostos e históricos ao que poderia ser encarados apenas como números. Utilizando o exemplo dado anteriormente, uma narrativa produzida partir da sobreposição dos mapas de criminalidade e de apreensão de drogas em uma determinada cidade poderia contar algumas histórias de vítimas que se encaixem no perfil resultante do cruzamento de informações, ou num dia a dia das comunidades identificadas como sujeitas à maior violência relacionada ao tráfico de drogas.

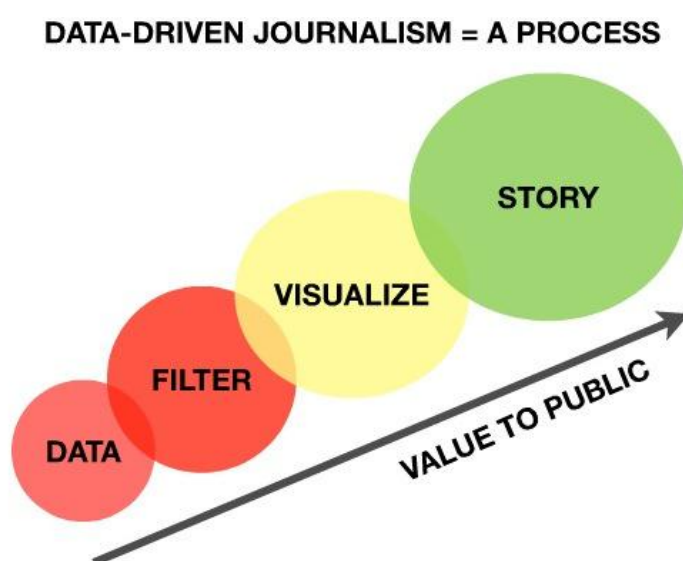


Fig. 7 – Representação do processo de desenvolvimento do jornalismo a partir de base de dados por LORENZ (*ibidem*, p. 13)

LORENZ (*ibidem*) identifica que o processo de produção de jornalismo a partir de base de dados, através das etapas citadas (dados, filtro ou algoritmo, visualização ou interface e narrativa) agrega valor ao conteúdo para o público ao qual é destinado. Para ele, estamos passando por um momento no qual o público deixará de ter que recolher pedaços de informações sobre o que os interessa que estão faltando e que terão que lidar com uma enxurrada de informação e a tentativa de filtrá-las para encontrar o que realmente os interessa. Na verdade, com a expansão dos dispositivos digitais de acesso às redes informatizadas e com a expansão do acesso em banda larga em boa parte do mundo, pode-se considerar tal cenário como virtualmente certo.

Sendo assim, faz-se necessário pensar que o que está sendo feito no jornalismo *online* a partir de bases de dados pode ser um bom indício das questões e necessidades aos quais o fazer jornalístico será submetido nos próximos anos. Não há subsídios suficientes para afirmar que esta prática ganhará forças e adeptos o suficiente para se tornar uma tendência ou estilo de jornalismo; seu histórico relacionado ao jornalismo de precisão e a RAC sugere que ela pode prosseguir como uma técnica limitada a pequenos grupos de jornalistas em busca de novas ferramentas metodológicas. O desafio de lidar com quantidades maiores de informações sendo disponibilizadas por entidades públicas e privadas ou mesmo sendo produzidas a partir de experiências de colaboração entre os indivíduos ligados às redes digitais pode fazer com que mais profissionais se aproximem dessa prática e seu uso cresça exponencialmente a partir dos trabalhos e experiências desses indivíduos.

## CONCLUSÃO

O jornalismo se encontra em um período de adaptação em relação às novas tecnologias de informação e comunicação. Podemos dividir essa adaptação em dois movimentos distintos, embora tenham ocorrido e ocorram de forma paralela em diversos momentos. Primeiramente, ocorreu uma rápida e progressiva adoção das tecnologias digitais na produção do jornalismo, como na adoção do computador para o processamento do texto e de imagens, a substituição de máquinas fotográficas analógicas por modelos digitais que alteraram a maneira de produzir e utilizar as imagens, a interligação dos computadores das redações e de praças através da chamada *intranet*, modificando processos de produção e de comunicação entre os profissionais, para citar apenas alguns exemplos.

O surgimento e disseminação da *internet* trouxe uma nova mudança, agora intimamente relacionada à matéria-prima da atividade jornalística. É claro que os processos de produção foram novamente alterados, como nas mudanças sofridas no processo de apuração jornalística que começou a ter que lidar com grandes volumes de informações que podiam ser rapidamente acessadas sem que se saísse da redação. Mas essa interligação de equipamentos digitais e, mais importante, das pessoas e instituições por trás desses equipamentos provocou mudanças no que podemos tomar como a principal matéria-prima da atividade jornalística: a informação. Toda a rede de relações de acesso, produção e distribuição de informações se alterou drasticamente com a expansão do acesso a equipamentos de comunicação digitais. Num momento, a imprensa perdeu a espécie de “monopólio” como veículo informativo capaz de atingir grandes públicos, já que qualquer indivíduo com uma informação de interesse poderia disseminar através da *internet* em canais de grande extensão de penetração, em qualidade bastante próxima da dita “profissional” - e por vezes superior. Muitos veículos de mídia sofrem com o descrédito do público e veem escapar, juntamente com a exclusividade da informação, o monopólio da credibilidade. Através de veículos como *blogs*, indivíduos estabelecem-se como fontes importantes de informação sobre os nichos sobre os quais se propõem dedicar, contando com muito mais crédito por parte de seus espectadores.

Dessa forma, a prática do jornalismo se encontra numa situação bastante peculiar. De um lado, encontra-se num ambiente extremamente favorável para quem trabalha com informação, já que há uma demanda por notícias e novidades de diversos tipos para atender os interesses de diversos grupos, sem falar nas novas ferramentas e meios para produzi-las. Por outro lado, a crescente facilidade de acesso e uso gerou, ao mesmo tempo, um cenário onde os veículos midiáticos tem que disputar a atenção e o tempo dos indivíduos com outros atores que não existiam até a muito pouco tempo, quando não tem que disputar espaço com a própria iniciativa de seu espectador tornar-se um produtor de informação. Nas palavras de LORENZ (2010) “o jornalismo está sob sítio. Os modelos de negócios dos jornais que funcionaram durante mais de 100 anos estão desabando.” (p. 11). Já não basta um bom texto para sustentar o jornalismo, pois há milhões de vozes que não são jornalistas por profissão ou vínculo institucional, mas que podem entender e relatar boas informações. É preciso achar um novo caminho, um caminho que torne a prática do jornalismo relevante novamente.

O uso de bases de dados para a produção de jornalismo nesse ambiente ainda pode ser considerada como uma especialidade. Um conjunto de práticas e técnicas desenvolvidas e utilizadas por uma parcela muito pequena de jornalistas e que exige uma formação ainda bastante específica em matérias consideradas diversas do ofício, como noções de matemática e estatística e programação de computadores. Formação esta que escapa do currículo da maioria das escolas de jornalismo e que raramente encontra mestres nas redações, reforçando o desenvolvimento individual do profissional que se dispõe a estudá-la.

SOARES (2011) chamou esses indivíduos de “elite cognitiva”, um pequeno grupo que se dedica a conseguir desenvolver o raciocínio e ferramentas necessárias para manipular estes dados e extrair deles informações que não foram e não seriam disponibilizadas ou mesmo descobertas de outras formas, já que muitas dessas informações escapam até mesmo das fontes dos dados interpretados. E é desse pequeno grupo, dessa prática pouco disseminada, que surge uma das mais interessantes propostas de formato e de “formas de fazer” que podem gerar um estilo próprio para esse campo ainda pouco compreendido que é o jornalismo

produzido para os meios digitais.

A prática do jornalismo a partir de base de dados públicos é algo que só pode ser realizado de forma plena em ambientes digitais e de rede; e se aproveita justamente das características desse ambiente para definir-se. As propriedades do ambiente digital – representação numérica dos conteúdos, modularidade, multimídia – são incorporados nessa linguagem, nessa forma de narrar fatos. O algoritmo informacional que é construído submete os dados a transformações e interpretações de forma dinâmica, de acordo com as solicitações e necessidades do leitor.

Ao construir um produto informacional que permite que seu público entre em contato com informações extraídas diretamente de dados primários fornecidos pela administração pública, através de uma interface que seja inteligível a ele, o jornalista permite que o indivíduo efetivamente se apodere daquela informação. E o faça de acordo com suas necessidades e não de acordo com qualquer tipo de edição que defina a prioridade de suas necessidades ou o filtro da fonte oficial que muitas vezes divulga as informações, mas esconde o acesso através de interpretações ou ângulos ditos “oficiais”.

Isso significa que a prática do jornalismo *online* a partir de bases de dados públicas pode ser, se realizado com esse compromisso, uma nova ferramenta de construção da cidadania. Se isso é importante para a sociedade, pode também representar uma renovação do próprio ofício do jornalismo, cada vez mais carente de respaldo público e de relevância.

Parece razoável também afirmar que a prática do jornalismo *online* a partir de bases de dados públicas parece ser hoje uma das mais bem-adaptadas às características e possibilidades dos meios digitais de comunicação e informação. E, como experiência, é importante também por trazer e estimular importantes contribuições como a incorporação de metodologias mais fundamentadas na lógica matemática, que permitem a interpretação mais correta de determinados fatos que escapam ou mesmo se distorcem se abordados pelas técnicas clássicas do

jornalismo; ampliam a caixa de ferramentas técnicas da profissão e obrigam aos seus profissionais repensarem práticas e buscarem novos conhecimentos.

Por tudo isso, essa pequena prática pode não chegar aos holofotes nas redações e escolas de jornalismo. Mas pode ser a origem de um muito bem-vindo sopro de renovação para esta *praxis* num momento onde ela se encontra mais desestabilizada desde sua criação.

#### 4. METODOLOGIA

A hipótese inicial que foi formulada para esta pesquisa é a de que a utilização de dados públicos disponibilizados em bases de dados digitais pelo jornalismo online permite a produção de informações jornalísticas relevantes a partir do relacionamento dinâmico destes dados, permitindo que a interpretação do significado dos dados seja mais acessível àqueles para os quais essas informações são produzidas. O gerenciamento dessas bases de dados e a compreensão e produção de interfaces que tornem esses dados compreensíveis para o leitor, ou seja, que os contextualizem e gerem efetivamente informação jornalística é uma tarefa e um desafio para os jornalistas que atuam nos ambientes digitais.

Para a experimentação da hipótese proposta, a decisão por apenas uma via metodológica seria um limitador ao trabalho e não daria conta das exigências do objeto a ser pesquisado. Desta forma, decidiu-se por utilizar uma múltipla abordagem metodológica às questões propostas.

Decidiu-se que o método qualitativo daria melhor conta tanto das exigências das análises a serem feitas para verificação da hipótese quanto pela própria natureza do objeto da pesquisa. Métodos quantitativos não serviriam para realizar uma exploração mais detalhada nas características da produção jornalística em ambientes digitais produzida a partir de bases de dados públicas. Além disso, considerou-se que o ainda pequeno número de trabalhos relevantes nesta área não justificaria a utilização desta metodologia.

Resolveu-se então que a metodologia que melhor atenderia as necessidades dessa pesquisa seria a realizar entrevistas com os responsáveis diretos pelos trabalhos jornalísticos selecionados com o objetivo de compreender melhor as práticas a partir do ponto de vista de quem as realiza. De acordo com os parâmetros propostos por ROSA e ARNOLDI (2008), foi utilizada a técnica de *entrevista especializada*, ou seja, aquela feita para um entrevistado que conheça bem o assunto a ser abordado e que permite que defina a situação, estruture seu relato e introduza suas noções do que considera relevante sobre o assunto, ao invés de

reservar esse papel somente ao entrevistador. Essa estrutura de entrevista foi escolhida já que se considera que os autores dos trabalhos selecionados possuem um conhecimento sobre o tema que justifica sua categorização como especialistas no assunto e cujas contribuições ao processo de entrevista pode ajudar a reunir informações importantes sobre o objeto desta pesquisa.

Foram estabelecidos então os critérios para selecionar os profissionais a serem entrevistados. Foram selecionados para esta pesquisa produtores de conteúdos jornalísticos *online* produzidos em português, que utilizassem quaisquer dados originalmente disponibilizados *online* por quaisquer instâncias governamentais (executivo, legislativo ou judiciário em esfera federal, estadual ou municipal) e que produzam interfaces destinadas ao acesso ou cruzamento desses dados. Prêmios nacionais ou internacionais obtidos a partir da produção de trabalhos que atendessem aos critérios relacionados anteriormente foram considerados diferenciais para a seleção, assim como relevância comprovada através de referências em reportagem jornalísticas realizadas devido a inovações relacionadas ao uso de bases de dados públicas para a produção de informação jornalística.

Foram selecionados ao fim dois profissionais para compor a base de entrevistas para essa pesquisa: Daniela B. Silva e Marcelo Soares.

Daniela Silva é jornalista, graduada pela Faculdade Cásper Líbero em 2007. Atuou na Editora Abril e Editora Moderna. Fez um curso de jornalismo multimídia pelo *Knight Center for Journalism in America*, programa de treinamentos ligado à Universidade do Texas. Graduou-se Mestre em Comunicação pela Faculdade Cásper Líbero em 2010, tendo defendido uma tese sobre a transparência governamental na esfera pública interconectada. Em 2009, fundou a Esfera junto com Pedro Markun, projeto destinado a trabalhar com comunicação, política e novas tecnologias, com o objetivo de realizar ações que estimulem a possibilidade de realização de um governo aberto a partir das possibilidades das tecnologias em rede.

Marcelo Soares é jornalista, graduado pela Universidade Federal do Rio

Grande do Sul em 2004. Atuou como repórter na Folha de S. Paulo. Foi um dos fundadores da Associação Brasileira de Jornalismo Investigativo (ABRAJI), onde foi gerente e instrutor de Reportagem Assistida por Computador (RAC). Foi pesquisador no “Transparência Brasil”, onde idealizou e executou o projeto Excelências, ganhador do 6º Prêmio Esso de Jornalismo na categoria “Melhor Contribuição à Imprensa” em 2006. Foi instrutor de RAC para diversos jornais e instituições em todo o Brasil, além de trabalhar como repórter *freelancer* para veículos nacionais e internacionais. Foi colunista de política na MTV entre 2009 e 2010. Marcelo retornou para Folha de S. Paulo em 2010, onde trabalhou nas editorias de Treinamento e Arte, na qual tentou produzir visualizações de informações. Atualmente, trabalha na criação de uma nova fase do *website* Bubot.com. É tradutor de histórias em quadrinhos para a Panini Brasil.

As entrevistas foram conduzidas através de roteiro semiaberto. A partir de perguntas-chave, o tema relevante da pesquisa foi introduzido para cada entrevistado. A partir das respostas dessas perguntas e do andamento geral da entrevista, outras perguntas foram feitas com o objetivo de conduzir o entrevistado a focar-se em conceitos ou opiniões a respeito do tema considerados importantes, ampliando a obtenção de informação nesses pontos. Foram estabelecidas uma série de perguntas para estabelecer o perfil profissional e acadêmico do entrevistado, com o objetivo de estabelecer a amplitude de conhecimento deste sobre o objeto desta pesquisa, justificando a escolha desta metodologia. O caráter aberto da entrevista também permitirá que o entrevistado, considerado de notável conhecimento sobre o tema, introduza informações que ainda não são conhecidas pelo pesquisador, ampliando a validade e contribuição da entrevista para a pesquisa.

As perguntas estabelecidas para o roteiro foram as seguintes:

- Sobre o projeto:
  - Como foi concebido o projeto? Qual era o objetivo inicial do projeto?
  - Quais bases de dados foram utilizadas para compô-lo?
  - Houve a necessidade de reconstruir as bases de dados utilizadas ou elas

foram utilizadas da forma que eram/são disponibilizadas pelo órgão público responsável por ela?

- Além das bases de dados públicas, usou-se alguma outra base de dados nesse projeto? Por quê?
- Quais foram as principais dificuldades encontradas durante a produção deste projeto?
- Qual foi a relação entre o resultado pretendido com o projeto e o alcançado? Atingiu-se o objetivo proposto, este não foi alcançado ou foi superado? Por quê?
- A partir desta experiência, quais são as principais responsabilidades, requisitos e desafios do jornalista ao produzir conteúdo para mídias digitais a partir de bases de dados?
- A partir desta experiência, qual é sua visão sobre o papel do uso de bases de dados no jornalismo online na produção jornalística nacional e internacional?

• Sobre o entrevistado:

- Qual é sua formação acadêmica e profissional?
- Quando e como começou seu trabalho com relação ao uso de bases de dados públicas para a produção de conteúdo jornalístico para mídias digitais?
- Quais conhecimentos teóricos e práticos foram exigidos de você nessa atividade? Como você os obteve/obtem? Quais são seus referenciais teóricos para a execução dos projetos de conteúdo jornalístico com o uso de base de dados?

O desenvolvimento da entrevista demonstrou que as perguntas específicas sobre o projeto não forneciam informações consideradas relevantes para o objetivo traçado para a pesquisa. Alterou-se o roteiro, então, para perguntas relacionadas com a experiência do entrevistado com o uso de bases de dados públicas no jornalismo e sobre a abertura de bases de dados públicas. O método de entrevista de roteiro semiaberto se mostrou eficaz nesse caso, pois permitiu que a entrevista fosse conduzida para as áreas de maior conhecimento e experiência de cada entrevistado. Ambas as entrevistas foram registradas em áudio; a entrevista com Daniela Silva foi registrada em vídeo.

As entrevistas foram decupadas e corrigidas. Elas foram a base para a identificação dos pontos mais importantes a serem pesquisados na etapa seguinte da pesquisa: a revisão bibliográfica. Os textos que já haviam sido reunidos durante a produção do projeto de pesquisa foram revisados e outros foram acrescentados a esta revisão de acordo com a necessidade apresentada tanto a partir da pesquisa quanto pelas leituras.

A redação e revisão do texto final da pesquisa foram realizadas na última etapa da pesquisa, a partir das anotações realizadas durante a revisão bibliográfica.

## BIBLIOGRAFIA

ANDERS CHYDENIUS FOUNDATION. *The World's First Freedom of Information Act*. Kokkola (Finlândia): Art-Print, 2006. Disponível em <http://www.scribd.com/doc/5885744/The-Worlds-First-Freedom-of-Information-Act-SwedenFinland-1766>. Acessado em 24 mar 2011.

ARNOLDI, Marlene Aparecida Gonzalez Colombo; ROSA, Maria Virgínia de Figueiredo Pereira do Couto. *A entrevista na pesquisa qualitativa: mecanismos para validação de resultados*. Belo Horizonte: Autêntica, 2008.

ASSOCIAÇÃO BRASILEIRA DE JORNALISMO INVESTIGATIVO (ABRAJI). *Curso de RAC – Reportagem com Auxílio de Computador ou Como transformar o computador no melhor amigo do jornalista*. Disponível em [http://www.arquivoz.com.br/arquivo/Curso\\_CAR\\_Abraji\\_rio.pdf](http://www.arquivoz.com.br/arquivo/Curso_CAR_Abraji_rio.pdf). Acessado em: 21 jun 2011.

BARBOSA, Suzana. *Jornalismo digital de terceira geração*. Covilhã, Portugal: Universidade da Beira Interior: 2007.

BARBOSA, Suzana. *Jornalismo Digital em Base de Dados (JDBD): Um paradigma para produtos jornalísticos digitais dinâmicos*. Dissertação (Doutorado em Comunicação e Cultura Contemporâneas). UFBA. Salvador: 2007.

BENNETTI, Marcia; LAGO, Claudia (orgs.). *Metodologia de pesquisa em jornalismo*. Petrópolis: Vozes, 2008.

BORGES, Juliano da Silva. *Webjornalismo: política e jornalismo em tempo real*. Rio de Janeiro: Apicuri, 2009.

BRASIL. Presidência da República. *Constituição República Federativa do Brasil de 1988*. Disponível em [http://www.planalto.gov.br/ccivil\\_03/constituicao/constitui%C3%A7ao.htm](http://www.planalto.gov.br/ccivil_03/constituicao/constitui%C3%A7ao.htm). Acessado em 23 mar 2011.

\_\_\_\_\_. \_\_\_\_\_. *Decreto Nº 7.185, de 27 de maio de 2010*. Disponível em [http://www.planalto.gov.br/ccivil\\_03/\\_Ato2007-2010/2010/Decreto/D7185.htm](http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2010/Decreto/D7185.htm). Acessado em 24 mar 2011.

\_\_\_\_\_. \_\_\_\_\_. *Lei Complementar Nº 101, de 04 de maio de 2000*. Disponível em [http://www.planalto.gov.br/ccivil\\_03/Leis/LCP/Lcp101.htm](http://www.planalto.gov.br/ccivil_03/Leis/LCP/Lcp101.htm). Acessado em 24 mar 2011.

\_\_\_\_\_. \_\_\_\_\_. *Lei Complementar Nº 131, de 27 de maio de 2009*. Disponível em [http://www.planalto.gov.br/ccivil\\_03/Leis/LCP/Lcp131.htm](http://www.planalto.gov.br/ccivil_03/Leis/LCP/Lcp131.htm). Acessado em 24 mar 2011.

CAIRO, Alberto. *Infografía 2.0*. Madrid, Espanha: Alamut, 2008.

CANADÁ. Presidente do Conselho do Tesouro do Canadá e pelo Ministro da Justiça e Procurador Geral do Canadá. Government of Canada's Access to Information Review Task Force. *Consultations of the Task Force with other Jurisdictions on their Access to Information Regimes*. Disponível em <http://www.atirtf-geai.gc.ca/consultation-sweden-e.html>. Acessado em 22 mar 2011.

CASTELLS, Manuel. *A sociedade em Rede – A era da Informação: Economia, Sociedade e Cultura. Vol. 1*. São Paulo: Paz e Terra, 1999.

COSTA, Rogério da. *A cultura digital*. São Paulo: Publifolha, 2008.

COX, Melisma. The development of computer-assisted reporting. Disponível em <http://com.miami.edu/car/cox00.pdf>. Acessado em: 22 jun 2011.

DEAK, André. *Novos jornalistas do Brasil: casos de processos emergentes do jornalismo na internet*. Dissertação (Mestrado em Ciências da Comunicação). USP. São Paulo: 2011.

ELMASRI, Ramez e NAVATHE, Shamkant. *Sistemas de Bancos de dados*. São Paulo: Pearson Addison Wesley, 2011.

#### DICIONÁRIO CALDAS AULETE DIGITAL

EUA. Departamento de Justiça dos Estados Unidos. *Freedom of Information Act*. Disponível em <http://www.foia.gov/>. Acessado em 23 mar 2011.

EUA. Departamento de Justiça dos Estados Unidos. *FOIA Guidance and Resources*. Disponível em <http://www.justice.gov/oip/foia-guidance-resources.html>. Acessado em 23 mar 2011.

EUA. Presidente dos Estados Unidos da América. *Transparency and Open Government*. Disponível em [http://www.whitehouse.gov/the\\_press\\_office/TransparencyandOpenGovernment/](http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment/). Acessado em 23 mar 2011.

EUROPEAN JOURNALISM CENTER (2010). *Data-driven journalism: What is there to learn?*. Disponível em <http://www.datadrivenjournalism.net>. Acessado em 29 jul 2011.

GOULARD, Caroline. *Database Journalism*. Disponível em <http://www.inaglobal.fr/en/numerique/article/database-journalism>. Acessado em: 21 jun 2011.

HOLOVATY, Adrian (2006a). *A fundamental way newspaper sites need to change*. Disponível em <http://www.holovaty.com/writing/fundamental-change/>. Acessado em 8 jul 2011.

\_\_\_\_\_ (2006b). *The programmer as journalist: a Q&A with Adrian Holovaty*. Disponível em <http://www.ojr.org/ojr/stories/060605niles/>. Acessado em 13 jul 2011.

\_\_\_\_\_ (2008). *In memory of chicagocrime.org*. Disponível em <http://www.holovaty.com/writing/chicagocrime.org-tribute/>. Acessado em 8 jul 2011.

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

MANOVICH, Lev. *The language of new media*. Cambridge, EUA: MIT Press, 2001.

MEYER, Phillip. *Precision Journalism*. Boston, EUA: Rowman & Littlefield Publishers, 2002.

PEW RESEARCH CENTER FOR THE PEOPLE & THE PRESS. Internet overtakes newspaper as news outlet. Disponível em <http://people-press.org/reports/pdf/479.pdf>. Acessado em 03 set 2010.

PULITZER PRIZES. Disponível em: <http://www.pulitzer.org>. Acessado em: 25 jun 2011.

SILVA, Daniela Bezerra da. *Transparência da esfera pública interconectada*. 2010. 114 p. Dissertação (Mestrado em Comunicação). Faculdade Cásper Líbero. São Paulo.

\_\_\_\_\_. *Transparência na esfera pública interconectada e dados governamentais abertos*. In: SILVEIRA, Sergio Amadeu (Org.) *Cidadania e redes digitais*. São Paulo: Comitê Gestor da Internet no Brasil – Maracá Educação e Tecnologias, 2010.

SOARES, Marcelo. *Jornalismo em Bancos de Dados*. São Paulo: 4 fev 2010. Aula ministrada na Escola de Comunicação e Artes da Universidade de São Paulo (ECA-USP).

SOARES, Marcelo. *Entrevista sobre jornalismo digital e bases de dados públicas*. São Paulo: 8 fev. 2011. Entrevista concedida a Mateus de Ornellas Rodrigues.

SUÉCIA. Parlamento da Suécia. *Freedom of Press Act*. Disponível em [http://www.riksdagen.se/templates/R\\_Page\\_\\_\\_\\_\\_6313.aspx](http://www.riksdagen.se/templates/R_Page_____6313.aspx). Acessado em 22 mar 2011.

UNIVERSITY OF MICHIGAN, INSTITUTE OF SOCIAL RESEARCH (2011). *Revealing the roots of a riot*. Disponível em <http://www.sampler.isr.umich.edu/2011/featured/revealing-the-roots-of-a-riot/>. Acessado em 22 jun 2011.

VALENTE, Ana Paola de Moraes Amorim. *Transparência e opacidade: o SIAFI no acesso à informação orçamentária*. São Paulo: Anablume; Belo Horizonte: FUMEC, 2004.