

John V. Pavlik

Professor do departamento de Jornalismo e Estudos de Mídia da Rutgers University.

**Antonio Brasil**

Professor doutor do departamento de Jornalismo da Universidade Federal de Santa Catarina (UFSC).



Big Data, Código Computacional e Arquivos de Notícias Televisivas: implicações dos avanços nos métodos de investigação audiovisual para a qualidade do jornalismo

Resumo

Organizações produtoras de notícias ao redor do mundo estão entrando na era do *Big Data*, dos algoritmos e dos códigos computacionais. Isso é mais do que uma reestruturação técnica das redações e dos métodos de reportagem. Essa confluência está inaugurando uma nova geração de métodos de pesquisa visual que traz implicações para valores e princípios, éticas e práticas que definem o jornalismo e seu papel na sociedade. Neste cenário de conexão e evolução da mídia digital, a televisão continua a ser fonte preeminente de informação para a maioria das pessoas. Filmes e vídeos de notícias televisivas, incluindo notícias em vídeo gravadas por cidadãos, estão entre os mais pervasivos e fundamentais registros da vida humana. Eles se apresentam como uma fonte primária, documentando eventos históricos específicos e cotidianos. Apesar do valor histórico, materiais noticiosos estão entre os mais ameaçados de extinção, são altamente subestimados e menos procurados em coleções de imagens em movimento. Arquivos de notícias têm sido fontes difíceis e caras de informações essenciais e de conhecimento para pesquisadores, historiadores ou mesmo para os próprios jornalistas. Além das questões financeiras e de direitos autorais, as ferramentas de pesquisa mais eficientes ainda estão em evolução e o acesso direto a grandes coleções *online* ainda é difícil e raro. Este artigo discute, portanto, o atual estado do jornalismo de dados e o potencial do código computacional para permitir que jornalistas investiguem grandes bases audiovisuais de dados pertencentes a arquivos televisivos noticiosos em diferentes países. **Palavras-chave:** Jornalismo; Big Data, Código Computacional; Algoritmos; Arquivos de Notícias Televisivas.

Abstract

News organizations around the world are entering an age of *Big Data*, algorithms and computer code. This is more than a technical restructuring of the newsroom and reporting methods. This confluence is ushering in a new generation of visual research methods that bring implications for the values and principles, ethics and practices that define journalism and its role in society. In this evolving and connected digital media landscape, television remains the world's pre-eminent information source for most people. Television news film and videotape, including citizen-recorded video of news events, are among the most pervasive and fundamental records of human life. They stand as primary source materials documenting specific historical and current events. Despite their historical value, news materials are among the most endangered, highly underestimated and less researched of moving image collections. News archives have been a difficult and expensive source of essential information and knowledge for researchers, historians, and journalists themselves. Financial questions, copyright issues, efficient research tools are still evolving and direct Internet access to large online collections is difficult and rare. This paper discusses the current state of data-driven journalism and the potential of computer code to enable journalists to investigate large audiovisual databases held in television news archives in different countries.

Keywords: Journalism; Big Data; Computer code; Algorithms; TV News Archives

Tradução_Daniel Melo Ribeiro e Leticia Xavier de Lemos Capanema.

DOSSIÊ
PRÁTICAS JORNALÍSTICAS

1. INTRODUÇÃO

O jornalismo sempre esteve em constante e intensa evolução. Superar desafios, dificuldades e crises com inovação e novas tecnologias está na natureza do negócio das notícias. Hoje em dia, empresas de jornalismo em diversas partes do mundo estão entrando em uma nova era digital. Os códigos computacionais e os algoritmos – essências do *Big Data* – apresentam novos desafios para a profissão do jornalista. O *Big Data* surge do grande conjunto de dados publicamente acessível em um momento de conectividade quase universal e disponível para análise e uso na cobertura jornalística. Códigos de computador referem-se a instruções computacionais que podem realizar tarefas no *Big Data*, incluindo a identificação de informações noticiosas relevantes, tais como fontes e tendências. Algoritmos são fórmulas matemáticas que jornalistas e programadores desenvolvem para usar em códigos computacionais a fim de extrair significado e identificar padrões no *Big Data*.

Apesar dos avanços em novas tecnologias digitais, da quase universal conectividade com a internet e do crescimento das mídias sociais, a televisão ainda é a fonte preeminente de informação para a maioria das pessoas. Noticiários são a principal, e algumas vezes a única, fonte de informação para a maioria das pessoas no mundo.

Cenas televisivas são também um importante registro de nossa própria cultura e nossa memória social. Em especial, as notícias locais retratam elaboradas imagens visuais e representam uma documentação histórica. Imagens em movimento possuem grande valor na educação e formação de nossas comunidades, capturando momentos representativos de nossa vida e de nossa sociedade e seus arquivos estão armazenados com informações valiosas e inexploradas.

Este estudo propõe o uso dos arquivos televisivos como uma nova e abundante fonte de *Big Data* para o jornalismo investigativo na era digital. Há poucos projetos extensivos e profundos de investigação jornalística audiovisual que exploram os arquivos de notícias televisivas como principal fonte de conte-

údo. Isso se justifica quando se considera as dificuldades legais de acesso a essas coleções e a carência de ferramentas de busca eficientes e especializadas em audiovisual. Ainda é bastante árduo, complicado e demorado obter autorizações para consultar boletins de notícias televisivas do passado. Esses programas de TV deveriam ser considerados como “documentos históricos primários” e deveriam portar um status especial de *copyright*, tendo em vista sua disponibilidade para investigação por jornalistas, pesquisadores ou cidadãos engajados.

É importante reiterar que telejornais são mais do que meros programas de televisão. Eles são também registros audiovisuais subestimados não-oficiais da nossa história recente e deveriam estar disponíveis em instituições públicas como bibliotecas ou arquivos, pelo menos para consulta e sem contratempos, para pesquisadores interessados. Eles deveriam estar acessíveis ao público em geral da mesma maneira que outras fontes importantes de informação, como livros ou jornais estão disponíveis. Não há motivo para considerar os telejornais de maneira diferente de qualquer outra fonte primária de informação histórica para futuros pesquisadores.

2. PESQUISA E ARQUIVOS DE NOTÍCIAS TELEVISIVAS

No passado, as restrições tecnológicas, a limitada disponibilidade de equipe técnica para auxiliar pesquisadores e os custos financeiros eram justificativas razoáveis para o acesso restrito aos arquivos de televisão em geral e aos noticiários em particular. Atualmente, com projetos extensivos de digitalização e a possibilidade de acesso direto *online* a vastas coleções de vídeos, tais justificativas não são mais relevantes ou aceitáveis. Pesquisas extensivas em amplos repositórios audiovisuais se tornaram muito mais frequentes, eficientes e descomplicadas em vários países como a França, os Estados Unidos e o Reino Unido.

Não obstante, em outros países, principalmente nos países em desenvolvimento, redes privadas de *broadcast* permanecem como os únicos proprietários de arquivos de televisão e eles tendem a negar

ou impor sérias restrições para acessar suas coleções de notícias televisivas. Essas restrições variam desde a necessidade de apresentar intenções explícitas de investigação por jornalistas, ou amplos projetos de pesquisa por acadêmicos, a condições técnicas e limitações de tempo. Essas justificativas para restringir, negar ou simplesmente não fornecer qualquer resposta oficial às requisições de acesso aos arquivos televisivos por pesquisadores geralmente fazem referência à falta de recursos técnicos ou financeiros, escassez de equipe especializada e outras sutis e inconclusivas alegações de ‘dificuldades’ para localizar o material requisitado.

De acordo com estudos recentes conduzidos por pesquisadores europeus, “pesquisar a história da televisão tem se tornado uma odisséia, o arquivo, uma caixa de Pandora e o arquivista, um porteiro kafkiano que nega ao pesquisador o acesso a tesouros escondidos da história da mídia - de maneira onipotente e intransigente¹”. Ao longo dos anos, os arquivos de televisão têm ganhado a reputação de serem fortalezas invulneráveis (cf. Oldenhege, 2000; Hecht, 2005; Ubois, 2005).

Apesar dessas dificuldades, há também amplas pesquisas tecnológicas sendo feitas em países como Estados Unidos e Japão para desenvolver ferramentas de pesquisa audiovisual mais eficientes, como *thesauri* especializados e programas de reconhecimento automático de rostos e cenas.

Isso representa um significativo auxílio, especialmente para investigações jornalísticas sensíveis ao tempo. Um dos principais desafios do jornalismo na era digital é conduzir pesquisas mais eficientes, especializadas, extensivas e profundas num curto período de tempo. E isso se torna ainda mais difícil para jornalistas que concentram suas investigações em fontes audiovisuais.

Nosso principal argumento é que os arquivos de televisão não devem restringir seu propósito a servir apenas suas próprias operações de *broadcast*

¹ Texto original: “researching the history of television becomes an odyssey, the archive a Pandora’s box and the archivist a Kafkaesque doorkeeper who denies the researcher access to hidden treasures of media history - almightily and uncompromisingly”. Fonte: http://www.researchingcommunication.eu/book9chapters/C17_Kramp1314.pdf

(Kramp, 2011b). Eles são importantes fontes de *Big Data* para o jornalismo investigativo, mas também são repositórios de documentos primários históricos para pesquisas acadêmicas e devem ter um status legal diferenciado e específico que considere o acesso a seus conteúdos.

Esta pesquisa pretende convencer jornalistas a usar os arquivos do noticiário televisivo de maneira mais regular, empregando de modo extensivo e eficiente as novas tecnologias digitais e ferramentas de pesquisa especializada, mas também pretende convencer instituições governamentais e redes de *broadcast* a garantir o acesso livre e gratuito aos arquivos televisivos. Arquivos esses que deveriam ter um status especial de *copyright* para garantir a jornalistas, pesquisadores e ao público em geral pelo menos a visualização (se não a cópia) desse conteúdo noticioso sem qualquer restrição.

3. QUADRO TEÓRICO DE REFERÊNCIA

Como referencial teórico para este estudo, adotamos a teoria *sense-making* de Dervin (1992), que afirma que a experiência é caracterizada pela descontinuidade. Esse conceito também pode ser visto como uma metáfora para o atual estado dos arquivos audiovisuais no qual múltiplos sistemas estão em uso e jornalistas e pesquisadores acadêmicos são obrigados a adotar novas estratégias e processos para garantir o acesso a eles. Além disso, o foco da teoria *sense-making* está mais no processo de resolução de problemas do que no sistema, fornecendo uma alternativa útil para interpretar a experiência de jornalistas e pesquisadores e para melhor compreender os passos que eles empreendem para cobrir as lacunas encontradas em suas atividades de processamento, pesquisa, descoberta e uso da informação. A teoria *sense-making* é apropriada pois concentra-se na importância da perspectiva, do conhecimento e das ações daquele que busca informações ao definir e potencialmente resolver problemas: “a maneira pela qual as pessoas veem suas lacunas estarão relacionadas com a maneira como elas tentam cobri-las e não com suas caracte-

terísticas pessoais e independentes das lacunas²” (Dervin, 1992).

Neste estudo, adotamos métodos qualitativos de pesquisa que incluem revisão da literatura, observação, monitoramento, exame, análise e interpretação de dados empíricos *online* e práticas correntes de pesquisa em arquivos audiovisuais em geral e arquivos de notícias televisivas em particular. O objetivo é apresentar, examinar e avaliar as práticas de pesquisa audiovisual, os padrões e as potenciais descobertas para jornalistas de investigações orientadas a dados. Enfatizamos a importância do acesso e compreensão dessas investigações como um fenômeno social que também requer observação, experiência e análise a fim de obter, para um jornalismo de qualidade, maior precisão e exatidão em detrimento de generalismos.

4. ARQUIVOS JORNALÍSTICOS E *BIG DATA*: UM EXAME CRÍTICO

Neste cenário em evolução, é fundamental compreender as origens, a importância e a relevância dos “arquivos” a fim de apreender seu potencial como uma nova fonte de *Big Data* para o jornalismo investigativo de televisão na sociedade contemporânea. De acordo com o historiador Jean Favier, “arquivos são um conjunto de documentos mantidos em ordem para provar direitos ou evidências de algumas atividades. Por metonímia, eles também designam o lugar (ou prédio ou instituição encarregada da conservação ou gestão) onde os documentos estão preservados” (Favier, 1959). A invenção do arquivo moderno enquanto lugar físico de coleta e também rastreamento da produção social de documentos de conhecimento histórico data da Revolução Francesa (Ricoeur, 2000).

Para o renomado historiador e filósofo francês Michel Foucault, “arquivos” significam práticas discursivas que envolvem sistemas, permitindo que declarações possam emergir como “eventos” e possam ser usadas ou ignoradas como “coisas”. Nesse

2 Texto original: “the ways in which people see their gaps will be related to the ways in which they try to bridge them and not to characteristics of persons independent of the gaps” (Dervin, 1992).

contexto, Foucault propõe chamar esses sistemas de declarações, coletivamente, de “arquivo”. Portanto, o arquivo não é somente uma coleção de textos que definem uma cultura, tampouco um conjunto de instituições que preservam textos. O arquivo é a “lei do que pode ser dito” e a lei de como o que é dito é transformado, usado, preservado etc. Assim, o arquivo pode ser definido como “um sistema geral de formulação e transformação de declarações³”. Ao invés de considerar o arquivo como um repositório de evidências não-mediadas sobre o passado, Foucault o vê como uma estrutura particular de poder na qual tipos particulares de documentos são mantidos em uma ordem particular, delimitando, assim, as possibilidades do que possa ser dito sobre os documentos e, portanto, sobre o próprio conhecimento.

O estudo pervasivo de arquivos modernos tem mostrado que a responsabilidade em manter seu conteúdo também traz um grande poder sobre uma parte significativa da herança cultural. Em seu ensaio *Mal d'Archive*, Jacques Derrida aponta a importância constitutiva dos arquivos para as atuais sociedades democráticas considerando as questões de poder e empoderamento. “Não há poder político sem controle do arquivo, se não da memória. A democratização efetiva pode sempre ser medida por um critério essencial: a participação e o acesso ao arquivo, sua constituição e sua interpretação” (Derrida, 1995).

Derrida está preocupado com a condição da verdade, e os arquivos se colocam exclusivamente pela verdade e pela evidência. Gestores de registros estão provavelmente cientes dos sistemas sócio-jurídicos que atribuem a “verdade” aos registros que eles gerenciam, assim como os arquivistas estão cientes das condições sob as quais seus objetos conservados podem ser considerados autênticos e devidamente comprobatórios. Cabe então a consideração dos pressupostos teóricos e epistemológicos por trás desses sistemas.

Arquivos e documentos, como registros de informação, também configuram nossa “memória coletiva”

3 “Um arquivo audiovisual é uma organização ou departamento de uma organização que possui uma permissão legal ou mandato para fornecer acesso a uma coleção de documentos audiovisuais e um patrimônio audiovisual através de coleta, gestão, preservação ou promoção” (Edmondson, 2004)

que é muito mais do que um agregado de memórias pessoais de indivíduos. Esses documentos, ou mapas de relevo, não são capazes de capturar o que uma nação inteira, por exemplo, coletivamente considera historicamente extraordinário ou banal. Para observar as “marcas” sociais do passado, precisamos, portanto, examinar o histórico social construído por todas as comunidades mnemônicas. Por isso devemos nos voltar aos locais inequivocamente sociais da memória (Zerubavel, 2003).

Documentam-se formatos de qualquer natureza (fotografia, jornais, vídeos, livros, mapas etc.) que assim constituem um sistema de suporte que concedem aos indivíduos o conhecimento, mesmo que através de elementos superficiais e subjetivos de sua história. Porém, apesar da importância do resgate de notas, esses documentos carregam um significado dotado de memória discursiva. Eles são montados em arquivos baseados em seleções aleatórias e, em alguns casos, em escolhas ideológicas. Ou seja, alguns são e devem ser lembrados enquanto outros são omitidos, censurados por aqueles que mantêm o poder de acumulá-los (Ferranezi & Romão, 2007).

Ademais, Henry Jenkins também denuncia os limites e controles impostos aos jornalistas investigativos e aos pesquisadores acadêmicos por aqueles responsáveis pela gestão dos arquivos televisivos. Sua crítica pode ser ainda mais relevante para aqueles dedicados às investigações “políticas” em coleções de notícias televisivas: “...isso determina o que é pesquisado e o que é relembado em diferentes períodos históricos. Isso reforça um preconceito particular do que foi a televisão em um momento e espaço particulares. E é muito difícil quebrar esse modelo realizando uma pesquisa arquivística, uma vez que a maioria das coisas que você talvez queira pesquisar pode não estar disponível⁴” (citado em Kamp, 2011).

Assim, a preservação e a disseminação de documentos históricos é necessária para recuperar e repensar o passado e para construir ou mesmo pre-

4 Texto original: “...it determines what is researched and what gets remembered from different historical periods. It reinforces a particular preconception of what television was at a particular time and place. And it is very difficult to break out that model by doing archiving research, because most of the stuff you might want to look at might not be available” (cited in Kamp, 2011).

ver o futuro. É essencial tornar tais registros disponíveis para os indivíduos, para que eles possam analisar seus conteúdos historicamente como parte da construção dos fatos, o que também sinaliza uma maneira de preservar a história contínua (Foucault, 2008). A partir disso, a acumulação dos documentos e das declarações geradas no tempo cria a necessidade de se organizar documentos em arquivos num fluxo histórico contínuo. Isso é fundamental para o jornalismo, no assim conhecido primeiro rascunho da história. Isso assume uma importância ainda mais especial na era do 24/7 e das notícias globalmente conectadas (Abramson, 2014)

5. ARQUIVOS AUDIOVISUAIS

A história do século XX foi escrita em imagens em movimento. Portanto, é importante considerar que os arquivos de larga escala preservam documentos audiovisuais que tem significado histórico e deveriam ser avaliados como projeções da nossa memória coletiva. Além disso, a emergência do arquivamento audiovisual como um campo especializado, separado dos arquivos gerais, é um desenvolvimento relativamente recente devido, em parte, à novidade das tecnologias das imagens em movimento. Porém, o valor ou a falta de valor atribuído aos documentos audiovisuais (Turner, 2009) também desempenha um papel no desenvolvimento de uma cultura de arquivamento do audiovisual. Ao longo dos últimos vinte anos, os arquivos audiovisuais estão passando por transformações significativas com a introdução da tecnologia digital (Lauwers, 2004; Ng, Rubin, & Van Malssen, 2010). Isso tem sido ainda mais relevante dentro das organizações televisivas de *broadcast*, que têm substituído sua infraestrutura pelo digital (Caldera-Serrano, 2008), permitindo que o conteúdo audiovisual seja copiado, armazenado ou mesmo em alguns casos específicos, tornando-os disponíveis *online*.

Arquivos audiovisuais tendem a se tornar mais versáteis, atendendo múltiplos usuários que podem acessar conteúdos simultaneamente e reutilizá-los em variadas situações. Essa nova configuração tecnológica cria uma excelente oportunidade para jornalistas produzirem histórias investigativas de maior

profundidade e garantirem o acesso geral para um público geral. Atualmente, no começo do século XXI, é possível dizer que os arquivos audiovisuais emergem como um importante campo de pesquisa.

A resolução 3442 da 18a. Conferência Geral da UNESCO, ocorrida em Paris em 1974, foi criada a fim de encorajar os estados membros a tomar medidas técnicas e legais para garantir a recuperação e o resgate de imagens em movimento de filmes e da televisão. Pelas mesmas razões, a Federação Internacional de Arquivos de Televisão (FIAT) foi fundada em 1977, contando com mais de 250 membros pelo mundo. Na comunidade em geral, há um reconhecimento crescente de que a televisão é um importante espelho e fonte da cultura. O documento da UNESCO “Recommendation for the Safeguarding and Preserving Moving Images”, publicado em 1980, afirma que “imagens em movimento são uma expressão da identidade cultural dos povos e, devido ao seu valor educacional, cultural, artístico, científico e histórico, formam uma parte integral da herança cultural de uma nação⁵” (idem).

Além disso, considerando essa herança cultural, François Jost reivindica que “os arquivos de televisão têm o potencial de escrever nossa história cultural e a comunicação ou as ciências da informação cometem um engano quando se limitam à atualidade. É hora de dar uma dimensão histórica aos arquivos de televisão. Sobretudo se quisermos que essas ciências desempenhem um papel na educação de nossos cidadãos” (Jost, 2005).

6. ARQUIVOS JORNALÍSTICOS COMO BIG DATA

Decifrar o *Big Data* e os algoritmos computacionais e permitir um maior acesso aos arquivos de notícias televisivas são um dos desafios mais recentes e importantes para o jornalismo de dados. Tais desenvolvimentos desafiam jornalistas e empresas do ramo a reconsiderar e reinventar a maneira como

5 Texto original: “moving images are an expression of the cultural identity of peoples, and because of their educational, cultural, artistic, scientific and historical value, form an integral part of a nation's cultural heritage”

eles coletam e produzem notícias, levando em conta, em especial, o crescimento massivo de arquivos de vídeo digital. Além disso, a evolução desses novos sistemas de produção de notícias permitirá mudanças na narrativa e na maneira de se envolver mais plenamente e de forma interativa os cidadãos em assuntos de importância pública. Fundamentalmente, essas forças estão conduzindo a uma transformação na qualidade do jornalismo na era *online* e digital. O contexto emerge como uma característica cada vez mais comum nas reportagens, conduzindo mídias inovadoras.

O *Big Data*⁶ levanta questões sobre privacidade, política, além de outros temas. Para o jornalismo e a mídia, o *Big Data* pode desempenhar papel-chave em narrativas (por exemplo, visualizações orientadas a dados), revelando padrões em eventos relacionados a notícias, dentre outros. Ferramentas para acessar e analisar o *Big Data* estão rapidamente se tornando essenciais para lidar com a qualidade das reportagens investigativas. Oportunidades de emprego em jornalismo e na mídia cada vez mais demandam habilidades em *storytelling* e reportagens baseadas em dados.

Obviamente, é importante posicionar o crescimento do *Big Data* em um contexto histórico. Por quase meio século, jornalistas se engajaram na mineração de dados (*data mining*) para criar reportagens sobre histórias. Já em 1968, Philip Meyer aplicou a análise de dados para criar uma reportagem sobre as revoltas contra o racismo em Detroit, ajudando o *Detroit Free Press* a ganhar o prêmio Pulitzer por seu jornalismo pioneiro. Isso alimentou uma era de reportagens investigativas baseadas em mineração de dados, conhecidas então por “Jornalismo de Precisão”, termo cunhado por Dr. Everette E. Dennis, reitor e CEO da Universidade Northwestern no Catar (NU-Q). No entanto, foi o relativamente recente e drástico crescimento da internet, da Web, das mídias sociais e móveis que inaugurou o fenômeno conhe-

6 O termo “*Big Data*” se refere ao conjunto massivo de dados que tem sido gerado pela sociedade digital e em rede, onde as informações sobre as pessoas, o planeta e muito mais estão sendo coletadas, analisadas e utilizadas em vastas quantidades e para efeitos de longo alcance. O *Big Data* frequentemente tem origem nas informações públicas disponíveis via mídias sociais, e suas implicações para a pesquisa e a educação são significativas.

cido como *Big Data* no século XXI. Suas consequências tocam quase todo aspecto da vida moderna.

Guiados pela convergência da computação, das redes e das tecnologias de coleta ubíqua de dados, o assim chamado *Big Data* está transformando um amplo espectro de atividades humanas. Dos esportes à ciência, do governo às mídias, as consequências do *Big Data* são muito diversificadas e intensas. Enquanto ainda não há uma concordância universal sobre o que seria o significado do *Big Data*, a noção de que grandes volumes de dados estão se tornando disponíveis de maneira crescente está sendo amplamente discutida e debatida, e suas implicações são profundas. Ferramentas de análises estão sendo empregadas para criar previsões baseadas no *Big Data* em um amplo espectro de campos, como a mídia, a política, os esportes e a saúde.

O movimento da codificação está criando incursões dramáticas no jornalismo contemporâneo, um setor que vem atravessando uma mudança radical em todo EUA e globalmente. Como os computadores, a informatização e os dados se tornaram fundamentais em praticamente todo aspecto da vida cotidiana, jornalistas e instituições jornalísticas começaram a reconhecer o papel central que a programação, ou codificação, na forma de algoritmos computadorizados ou instruções digitais, irá desempenhar nas reportagens e nas narrativas.

A urgência é particularmente aguda nesse domínio que acabou conhecido como *Big Data*, ou conjunto de dados que frequentemente são medidos em bilhões (ou mais) de bits de informação, muitas vezes coletados não somente das mídias sociais, mas também do governo e dos registros transacionais dos grandes negócios. Este último tem sido um pilar para o jornalismo investigativo desde, pelo menos, os anos 1980, embora as reportagens auxiliadas por computador tenham nascido nos anos 1960. Frequentemente, o *Big Data* emerge do mundo das mídias sociais, onde milhões ou mesmo bilhões de cidadãos da comunidade, da nação ou do mundo publicam mensagens para o consumo, consideração e comentário público. Muitas vezes, pesquisadores e repórteres percebem o *Big Data* como fontes de histórias, referências e tendências.

A confluência dos dados e algoritmos executados pelos códigos computacionais e a conectividade global em rede apresentam implicações transformadoras para o jornalismo na era da tecnologia digital e *online*. Para criar suas reportagens, os jornalistas confiaram, durante a maior parte do século passado, em entrevistas, conferências de imprensa e observações em primeira mão de eventos noticiosos, suplementados por informações coletadas em documentos e outros registros de agências governamentais, grandes empresas e outras instituições.

O crescimento da internet e, especialmente, da conectividade de praticamente todos e tudo à internet começou a mudar os fundamentos de como os jornalistas coletam as notícias, sem mencionar como eles contam histórias e as entregam ao público. Por sua vez, o público se torna altamente engajado na arena das mídias sociais, especialmente estimulada pela esmagadora adoção pública dos dispositivos móveis tais como os *smartphones* e *tablets* conectados à internet. Jornalistas têm sido confrontados por pelo menos duas novas realidades.

A primeira é que o público não é mais caracterizado pela passividade, quando se considera o mundo da mídia, especialmente as notícias. Ao invés disso, o público é, muitas vezes, ativamente engajado em ajudar nas notícias de última hora, também as retransmitindo ou “retuitando”. Os então chamados jornalistas ou repórteres cidadãos atuam como um quinto estado, complementar ao quarto estado dos jornalistas profissionais.

A dificuldade para o jornalismo é que o volume de dados é tão massivo e potencialmente dinâmico que uma ação humana individual ou mesmo um time de duas ou três pessoas (o *modus operandi* padrão do antigo jornalismo) não consegue, de maneira prática ou eficiente, em qualquer período estreito de tempo, como o *deadline* de um repórter, esperar destilar padrões nesses bilhões de *Big Data*. Consequentemente, para utilizar o *Big Data*, jornalistas precisam criar e usar códigos computacionais que implementam algoritmos de resolução de problemas para classificar os dados. O código frequentemente toma a forma de instruções nas quais, se uma certa condição é atendida, instruções adicionais

dirão ao computador o que fazer com os dados, possivelmente contando ou executando outros cálculos matemáticos que revelam padrões nos dados ou exceções gerais em padrões ou tendências.

No final dos anos 1960 e início dos anos 1970, uma versão mais antiga de reportagens criadas por computador e orientadas por dados foi classificada como “Jornalismo de Precisão” (Meyer, 1971). Uma razão para essa classificação foi que, no passado, os métodos jornalísticos, embora reveladores, eram muitas vezes anedóticos e sujeitos a caprichos sobre os eventos correntes. As notícias eram interrompidas e moviam-se de um tópico a outro sem que os jornalistas criassem conexões com tendências mais amplas ou situassem eventos noticiosos em um contexto histórico, sociológico ou econômico mais amplo. As reportagens não eram sistematizadas e careciam de precisão. Os dados que seriam analisados pelos repórteres em planilhas ou em outras ferramentas estatísticas adaptadas para as ciências sociais, permitiram aos jornalistas situarem questões em desenvolvimento de maneira mais contextualizada, ajudando a fornecer explicações e a criar maior sentido para as notícias.

O surgimento do *Big Data* e o uso correspondente dos algoritmos implementados através de códigos computacionais para interpretar os conjuntos massivos de números começaram a transformar não somente as reportagens investigativas ocasionais mas o jornalismo diário. Como exemplo dessa tendência para o mercado de notícias, a Associated Press anunciou que irá expandir seus programas de histórias automatizadas. Lou Ferrara, vice-presidente da AP e editor-chefe que orienta notícias de mercado, explica que “parte do sucesso dos 170 anos de história da AP tem sido a adaptação e a inovação em novos formatos de mídia como parte de nossa própria história”⁷.

7 Texto original: “part of the success of the 170-year history of AP has been adapting and innovating in new media formats as part of their own history”. Fonte: <https://www.washingtonpost.com/news/innovations/wp/2015/01/29/associated-press-looks-to-expand-its-automated-stories-program-following-successful-launch/>

7. ARQUIVOS DE NOTÍCIAS TELEVISIVAS⁸ COMO UMA IMPORTANTE FONTE DE *BIG DATA* PARA O JORNALISMO: CENÁRIO ATUAL

A televisão existe há aproximadamente 60 anos. A longevidade da televisão despertou um reconhecimento de sua importância para a herança cultural das nações, provocando o estabelecimento de instituições de arquivos. O fornecimento do acesso a esses arquivos é uma justificativa política para o financiamento público destinado ao arquivamento da herança audiovisual e também tem sido um objetivo compartilhado na comunidade ligada aos arquivos. (Yakel and Kim 2005).

É importante enfatizar que os arquivos de notícias televisivas e o jornalismo investigativo na era digital se apresentam como oportunidades de reinvenção dos métodos de produção investigativa que demandam busca e edição de arquivos de notícias em vídeo e em outras fontes. Durante a última década, os arquivos audiovisuais de televisão sofreram mudanças significativas em resposta à introdução das mídias digitais e dos sistemas de produção digital, particularmente em relação às notícias televisivas. Há pesquisas significativas sobre como o uso da tecnologia digital afeta outros grupos de usuários no setor de difusão televisiva, mas há poucas pesquisas que focalizam o trabalho dos jornalistas⁹.

Dezenas de milhares de horas de material de difusão televisiva juntamente com os registros do catálogo de muitos dos principais arquivos de televisão estão, agora, disponíveis *online*, mas as ferramentas de pesquisa ainda estão em evolução e o acesso direto pela internet a grandes coleções é raro. Arquivos de notícias têm sido uma fonte difícil e cara de conheci-

8 Arquivos de notícias televisivas - é uma área especializada de uma biblioteca audiovisual mais ampla. É geralmente responsável por coletar, organizar e fornecer acesso ao conteúdo gerado pelo departamento de notícias. Os arquivos de notícia fornecem serviços, em primeiro lugar, mas não exclusivamente, para o departamento de notícias e, devido à sua natureza única, possuem demandas que diferem de outros arquivos audiovisuais. Fonte: <http://researcharchive.vuw.ac.nz/xmlui/bitstream/handle/10063/2235/thesis.pdf?sequence=2>

9 <http://besser.tsoa.nyu.edu/howard/Papers/tv_archive.pdf>, <<http://researcharchive.vuw.ac.nz/xmlui/bitstream/handle/10063/2235/thesis.pdf?sequence=2>>

mento e de informação essencial para pesquisadores, historiadores e para os próprios jornalistas.

8. ARQUIVOS DE NOTÍCIAS DE TELEVISÃO: TENDÊNCIAS ATUAIS

Sendo importante não somente para retratar eventos atuais, as notícias são muito versáteis e muitas vezes são usadas numa variedade de outros gêneros: “notícias e programação de atualidades possuem o mais alto potencial de significação em relação a qualquer outro tipo de produção televisiva. Difusões de notícias televisivas incluem reportagens sobre *hard news* tais como (...) política e condições econômicas. Elas também tratam de histórias sobre personalidades proeminentes, eventos culturais e estilo de vida¹⁰” (Bergeron, 1986). Mais do que qualquer outro gênero, as notícias reúnem a dupla necessidade de documentar fenômenos históricos e culturais, bem como fornecer uma fonte confiável para ser usada em outras produções. Portanto, a seleção de valores e os processos adotados por arquivistas de audiovisual que trabalham em bibliotecas de notícias devem refletir as necessidades da organização, como também a comunidade mais ampla, como descrito por um bibliotecário de notícias: “(...) muitos de nós estão sempre cientes da responsabilidade em manter tanto um registro histórico do emissor como também um registro histórico público mais amplo da vida nacional tal como ela é refletida na televisão¹¹” (Muller, 1994).

O jornalismo de dados quebra as barreiras das notícias e estimula jornalistas a desenvolverem novas ferramentas investigativas. Enquanto isso, repórteres estão voltando suas atenções para o uso de novos e poderosos bancos de dados. Um desses imensos repositórios de informação são os arquivos de notícias televisivas. Mas, de acordo com Davidson S. e

10 Texto original: “News and current affairs programming have the highest potential significance of any type of television production. Television news broadcasts contain reports on ‘hard’ news such as (...) politics and economic conditions. They also feature stories on prominent personalities, cultural events, and lifestyles” (Bergeron, 1986).

11 “(...) many of us are always aware of a responsibility to maintain an historical record of the broadcaster as well as maintaining a wider historical public record of national life as reflected by television” (Muller, 1994).

Lukow G.: “Filmes e vídeos de notícias televisivas estão entre os registros mais pervasivos e fundamentais da vida norte-americana. Eles se apresentam como uma fonte primária, documentando eventos históricos específicos e que, ao mesmo tempo, revelam os ânimos, crenças, modas e atitudes mais intangíveis de nosso tempo¹².” Eddie Richmond, Presidente da Associação dos Arquivistas das Imagens em Movimento, acrescenta: “Apesar do seu valor histórico, as notícias estão entre as mais ameaçadas, altamente subestimadas e menos pesquisadas das imagens em movimento.¹³”

9. CARACTERÍSTICAS DO MÉTODO DO JORNALISMO INVESTIGATIVO AUDIOVISUAL

O século XXI não é somente uma “era da informação”. É um século visual. De acordo com o professor Mitchell Stephens, “o vídeo não somente preserva a memória, mas também organiza fatos¹⁴”. Nesse contexto, é também possível afirmar que a comunicação requer pensamento visual para o jornalismo visual e uma narrativa também visual.

Neste ponto, é essencial enfatizar que a investigação jornalística requer não somente “precisão”. É também “sensível ao tempo”. Ela requer velocidade e acesso amigável a fontes vastas e complexas como os arquivos de notícias televisivas. Porém, quase todos os aspectos da televisão e do vídeo hoje estão em transição, criando novos paradigmas tanto no *Big Data* quanto nos arquivos de notícias televisivas.

A convergência do jornalismo de dados e dos arquivos de notícias de televisão possui um imenso

12 Texto original: “Television news film and videotape are among the most pervasive and fundamental records of American life. They stand as primary source materials documenting specific historical events while at the same time revealing the more intangible moods, beliefs, fashions and attitudes of our times”. Fonte: <http://www.loc.gov/film/tvstudy.html>

13 Texto original: “Despite their historical value, news materials are among the most endangered, highly underestimated and less researched of our moving images”. Fonte: <http://www.amianet.org/sites/all/files/Awards%20-%20Leab.pdf>

14 Texto original: “Video not only preserves memory but organizes facts”. Fonte: Stephens, M., *The Rise of the Image, The Fall of the Word*, Oxford University Press, N.Y. 1998.

potencial de investigação de histórias de notícias audiovisuais. Mas essa convergência depende de muitos fatores, incluindo o desenvolvimento de ferramentas inovadoras para melhorar a possibilidade de desenvolvimento de novos e específicos recursos audiovisuais para análise comparativa da cobertura de notícias no âmbito nacional e internacional.

10. ESTUDOS DE CASO

Para explorar o emergente impacto do *Big Data*, dos algoritmos e da codificação na investigação jornalística usando arquivos de notícias televisivas junto a métodos de pesquisa audiovisual, escolhemos três das maiores e mais importantes instituições de arquivos de notícias no mundo. São elas a *News Archive* da Universidade Vanderbilt, nos Estados Unidos, os arquivos da BBC, no Reino Unido, e o *Institut National de l'Audiovisuel* (INA), na França. Cada uma dessas instituições foram consideradas proeminentes e essenciais, uma vez que contêm extensas coleções digitais, empregam práticas comuns para avaliar, registrar e catalogar material noticioso. Porém, adotam diferentes estratégias em relação ao acesso aos seus conteúdos. Estes estudos de caso certamente oferecem uma visão geral sobre os potenciais problemas e expectativas para futuros jornalistas orientados a dados e pesquisadores que decidirem pelo uso de arquivos de notícias televisivas.

A Universidade Vanderbilt, uma instituição educacional privada localizada em Nashville, no estado do Tennessee nos Estados Unidos, mantém o amplamente utilizado *Vanderbilt Television News Archive*. É considerado o arquivo mais extenso e completo de notícias televisivas no mundo¹⁵. Transmissões datadas desde 5 de Agosto de 1968 até os dias atuais¹⁶ são registradas e disponibilizadas. Os arquivos completos de vídeo dos Estados Unidos consistem, primordialmente, de noticiários e programas especiais de notícias das principais redes norte-americanas. O arquivo Vanderbilt possui mais de 850.000 trechos de notícias e cerca de 30 mil horas de conteúdo registrado e suas respectivas informações complementares.

15 <http://tvnews.vanderbilt.edu/?SID=20140813343933105>

16 18 de Agosto de 2014.

Em termos do acesso digital, é fácil navegar em sua plataforma disponível na internet, assumindo um conhecimento prévio da língua inglesa. Todos os documentos estão disponíveis ao público. Interessados podem requisitar cópias do material em DVD ou VHS. Porém, o material é entregue como um empréstimo. Isso ocorre em razão do respeito aos direitos de propriedade intelectual, ou *copyright*, dos documentos. Para fazer uma demanda de empréstimo, é necessário somente que o solicitante registre-se no site. Qualquer pessoa, não importando sua área profissional, pode realizar um pedido.

Outro sistema de arquivamento audiovisual pertence à *British Broadcasting Corporation* (BBC), que atualmente é a maior empresa de difusão no mundo. Formada em 1927, a BBC iniciou as transmissões de televisão em 1932, durante a era experimental do *broadcast*. Os primeiros programas produzidos pela BBC ainda são muito importantes e relevantes para a pesquisa do desenvolvimento da televisão britânica.

Estabelecido em 1948, o arquivo da BBC foi um dos primeiros a conservar o *broadcast*. Embora ele tenha existido primordialmente para reutilização dos programas, o arquivo da BBC teve um departamento interno de preservação que assegurou que os itens estivessem adequados para reuso e ativamente digitalizou os materiais para o fácil acesso e para a preservação a longo prazo das matrizes. Dentre materiais de televisão e de rádio, o arquivo atualmente contém aproximadamente 4 milhões de itens físicos. Desses 4 milhões de itens, há 600.000 horas de conteúdo televisivo¹⁷.

Registros anteriores a essa data estão, de maneira geral, menos preservados; o processo de gravação em fita magnética foi originalmente inventado em 1947, enquanto que a gravação em fitas de vídeo foi gradualmente introduzida a partir do final dos anos 1950. Mas, devido ao custo das fitas, as gravações eram consideradas somente para produções, tendo seus registros apagados subsequentemente.

Atualmente, a maioria dos programas são mantidos, incluindo notícias, entretenimento, dramati-

17 Lee, Adam. "The BBC Television Archive." BBC Archive. <http://www.bbc.co.uk/archive/tv_archive.shtml?chapter=6>.

zações e uma seleção de outros programas de longa duração, como os *quiz shows*. O material restante do arquivo televisivo é oferecido ao *British Film Institute* antes de ser descartado. Nos arquivos da *British Broadcasting Corporation*, há cerca de 4 milhões de itens físicos para a TV e para o rádio. Isso equivale a 600.000 horas de conteúdo de TV e cerca de 350.000 horas de rádio. Há também um arquivo de novas mídias, que mantém o registro do conteúdo que vai para os sites da BBC. A emissora dedicou muito tempo e esforço na preservação de seu arquivo que, recentemente, está sendo digitalizado. O conteúdo tem sido preservado em grandes arquivos digitais, tal como as pessoas mantêm suas próprias fotos ou arquivos de Word, que são guardados em vastos "armazéns de dados" (*data warehouses*). A questão mais fantástica sobre a digitalização é permitir que os arquivos se tornem muito mais flexíveis.

Atualmente, o projeto *BBC Genome* está criando um catálogo online completo de todos os programas da BBC para que as pessoas possam descobri-los em suas coleções. Eles estão trabalhando sob o princípio de que "o acesso público completo ou quase completo aos arquivos é não somente possível, como também é o objetivo final¹⁸". Situado no coração do reprojeto BBC online, o *BBC Genome* é o primeiro passo em direção a esse objetivo. Ele fornece uma linha do tempo desde a fundação da BBC em 1922 e também disponibiliza detalhes dos programas, canais e serviços que são mapeados nessa *timeline*, trazendo a história da BBC à vida.

O *BBC Genome* armazena informações sobre os programas de diferentes maneiras, cada uma designada a apoiar uma tarefa específica ou função. Mas nenhuma dessas tarefas é abrangente, tanto no acesso público quanto na pesquisa. O objetivo é assegurar que o histórico de transmissões *broadcast* da empresa, desde 1922, torne-se e permaneça ativo para o público.

O arquivo da BBC de Londres inclui principalmente trechos e programas completos desde notícias, atualidades e outros programas regionais. Materiais de direitos autorais foram incluídos em

18 Texto original: "full or near-full public access to archives is both achievable and the right ultimate goal".

um programa de exceção de acordo com a seção 30 do *Designs & Patents Act*, de 1988, que diz respeito ao acesso legal a propriedade intelectual, afirmando que há "proibidade para o propósito de reportar eventos atuais: parte do conteúdo do programa considerado para uso em um arquivo *online* irá conter material não esclarecido que foi originalmente incluído no programa com o propósito de reportar eventos atuais sob a proibidade das disposições da Lei de Direitos Autorais. Porém, o uso subsequente desse conteúdo, como por exemplo, torná-lo disponível em um arquivo *online*, não será considerado uso para o propósito de 'reportar eventos atuais'¹⁹". Consequentemente, há um risco de reclamação dos detentores dos direitos sob o uso do conteúdo em um arquivo *online* sem maiores esclarecimentos sobre o que está sendo empreendido. Para manter o risco em um nível baixo e administrável, é importante que o material disponível não possa ser baixado ou usado de outra forma (que não seja para pesquisa pessoal) e que o material permaneça dentro do contexto de uso original. Adicionalmente, o guia da BBC para pesquisadores deixa claro que o conteúdo de suas coleções de jornalismo é um registro de reportagens em um dado período e tal abordagem ressalta a necessidade de que o contexto em torno do uso permaneça²⁰.

A BBC está também conduzindo um projeto piloto utilizando boletins de notícias que começou no outono de 2011, examinando as questões que envolvem a publicação de notícias *online*. De uma perspectiva dos direitos de quem contribui, essas questões contemplam: a) a inclusão de cenas de esportes registradas por terceiros, b) o uso de exceção de proibidade na lei de direitos autorais para o propósito de reportar eventos atuais ou para crítica e comentários, restrições legais em torno do conteúdo, c) uso de material sensível e restrito, como obituários.

19 Texto original: "Fair dealing for the purposes of reporting current events: some of the program content considered for use in an online archive will contain uncleared material that was originally included in the program for the purpose of reporting current events, under the fair dealing provisions of the Copyright Act. However, the subsequent use of this content, for example making it available within an online archive collection, is unlikely to be considered use 'for the purpose of reporting current events'".

20 Guide To The BBC's Archives 2012 https://dl.dropboxusercontent.com/content_link/K63YHoyHV0woUkLxaXyKb8v97FEmKDr-BWBEPQd0e3a5JczbmOcyzPTerZlpkURAJ?dl=1

O terceiro exemplo é o *Institut National de l'Audiovisuel (INA)*, na França, que é o repositório central dos arquivos audiovisuais da televisão e do rádio franceses. Fundado em 1975 e aberto ao público desde 1998, a coleção se caracteriza por mais de centenas de milhares de documentos audiovisuais (incluindo programas históricos e de cunho jornalístico), totalizando 25 mil horas. Desde 2006, o INA abriu versões digitais para acesso gratuito, com arquivos escaneados e pesquisas *online*. O arquivo coleta e mantém o conteúdo de 17 estações de rádio e de 45 estações locais de televisão, incluindo programação de outros países (Hoog, 2006).

O INA coleta um total de aproximadamente 300.000 horas de televisão e rádio anualmente e a coleção é composta de mais de 2.5 milhões de horas de televisão e de rádio. A coleção data de mais de 60 anos e mantém algumas das primeiras e experimentais transmissões francesas. Através do apoio do governo e de seu próprio sucesso comercial, o INA se desenvolveu como um dos arquivos mais avançados no mundo. O Instituto tem trabalhado com materiais digitais há algum tempo e possui uma política local para a captura de sons e imagens em movimento para preservação digital de longo prazo. Em 2002, como parte de sua operação mandatária de depósito, o INA começou a capturar o sinal de 12 canais a cabo e por satélite para preservação.

Talvez a iniciativa digital mais conhecida do INA seja seu massivo projeto de “proteção e digitalização”. Iniciado em 1999, esse projeto resultou na digitalização de mais de 200.000 horas de materiais, principalmente conteúdo radiofônico e televisivo, bem como um *website* através do qual o conteúdo pode ser acessado. O website fornece a maior parte do conteúdo de graça como um serviço para o público. Aproximadamente 20% do conteúdo foi considerado pelo Estado como sendo um conteúdo criativo único por meio do qual o INA precisa remunerar seu criador. A taxa para esse conteúdo, porém, permanece razoavelmente baixa. O INA espera que “todo o arquivo ameaçado possa ser processado e salvo” até 2015²¹.

21 <http://www.institut-national-audiovisuel.fr/nous-connaitre/entreprise/chiffres-cles.html>

Centenas de programas televisivos, incluindo boletins de notícias, são adicionados diariamente como parte dos arquivos audiovisuais do INA. Desse total, 80% do conteúdo está disponível *online* gratuitamente. O pesquisador ou o público em geral pode simplesmente digitar no campo de busca o termo desejado para começar a assistir. Os outros 20% podem também ser acessados *online*, mas somente os primeiros minutos. Nesse último caso, para acessar o conteúdo completo, o usuário precisa adquirí-lo na loja do arquivo. Lá, o usuário poderá encontrar todos os vídeos e áudios disponíveis em diferentes mídias tais como CD ou DVD. De acordo com informações do próprio site, as taxas são aplicadas a fim de remunerar os detentores dos direitos autorais do arquivo, bem como para financiar a própria coleção. Para acessar o conteúdo disponível na loja do arquivo é também necessário completar um cadastro no *site*. Assim que o cadastro for completado, um *e-mail* é enviado ao usuário que, após confirmar seu endereço eletrônico, pode aproveitar o conteúdo livre de encargos.

Além disso, os arquivos de televisão podem também ampliar e enriquecer seu escopo coletando conteúdo “social” ligado aos programas transmitidos. O novo projeto digital do INA é denominado “Arquivo de Broadcast Aumentado” e consiste em registrar, indexar e catalogar não somente transmissões de rádio e de televisão mas também 12.519 websites, 5000 contas do Twitter, 26 rádios *online* no formato *streaming* e 8M de vídeos hospedados exclusivamente *online*. Esse é um grande passo em direção a um sistema mais completo de arquivamento digital de um dos acervos digitais mais relevantes, preparando a instituição francesa para os próximos desafios da comunicação na internet. Isso irá, certamente, acrescentar novas perspectivas de busca para jornalistas investigativos ou pesquisadores acadêmicos.

Tal como nos casos americano e inglês, o *site* francês é também intuitivo, com fácil navegação, demandando um conhecimento básico da língua francesa. Na página de ajuda, há instruções passo-a-passo sobre como completar o cadastro, conduzir buscas e acessar outras informações para o uso dos arquivos. Usuários são também informados sobre os

requisitos técnicos para assistir o conteúdo *online* e *offline*, caso haja interesse na aquisição de conteúdo para uso em documentários ou em outros projetos.

Após essa breve descrição e análise sobre as três instituições estudadas, é possível, agora, listar e enfatizar alguns pontos semelhantes de desempenho que são essenciais para o jornalismo investigativo orientado a dados e para pesquisadores acadêmicos que usam arquivos de notícias televisivas: a disponibilidade do arquivo *online*; a garantia do acesso aos usuários interessados; a oferta gratuita de documentos desejados ou a partir do pagamento de taxas.

Os estudos de caso também revelam os diferentes status em relação aos recursos disponíveis, a extensão do conteúdo e as estratégias de acesso às coleções audiovisuais em geral e ao material de notícias, em particular. Ainda há muita dificuldade para tornar esses conteúdos disponíveis para jornalistas, para pesquisadores ou para o público em geral. Mas, de acordo com estudos mais específicos, já é possível notar e avaliar a contribuição do material de arquivo não somente para jornalistas ou pesquisas, mas também para produtores de notícia de televisão: “a documentação audiovisual enriquece e completa as notícias, facilitando a compreensão da audiência²²” (Agirreazaldegí, 2008). Nesse mesmo contexto, é também notável uma pressão crescente sobre as empresas de *broadcast* para oferecer programação para variadas plataformas digitais, bem como a importância dos benefícios econômicos de se capturar e preservar material para reuso, mudando a percepção das empresas de audiovisual que passam a considerar esse material como um ativo e não um passivo (Murphy, 1997).

11. PESQUISA E RECUPERAÇÃO: NOTÍCIAS E ATUALIDADES

O principal grupo de usuários que confia plenamente nos arquivos de televisão são os jornalistas e produtores. Mas, como usar arquivos de notícia televisivas como uma importante e relevante fonte de *Big*

22 Texto original: “audiovisual documentation [library content] enriches and completes the news, facilitating the audiences understanding” (Agirreazaldegí, 2008).

Data no jornalismo investigativo? Como encontrar a imagem adequada em arquivos de notícias de TV?

É importante enfatizar que a busca nos arquivos de notícias televisivas demandam: a) planejamento, b) observação, c) indexação/classificação, d) registro, e) edição e f) análise do conteúdo. A elaboração de ferramentas automatizadas ou amigáveis de recuperação e busca de vídeos para o jornalismo investigativo orientado a dados pode, algumas vezes, ser um processo demorado e caro nos arquivos de notícias televisivas. Jornalistas investigativos tendem a demandar a localização de imagens específicas em curto prazo, e isso significa que eles confiam plenamente nas descrições detalhadas dos catálogos. Além do mais, a introdução da tecnologia digital na produção de notícias alterou o papel do jornalista, que agora assume tarefas de edição, incluindo a localização de material bibliográfico e a produção de material para uma crescente variedade de canais e plataformas.

Espera-se que os usuários possam melhor explorar as ferramentas de pesquisa. Argumentamos aqui que o processo poderá ser aprimorado com a inclusão de imagens nos registros dos catálogos. Ainda há uma compreensão limitada entre usuários sobre as estratégias de busca avançada. Cada vez mais, a falta de conhecimento sobre uma coleção em particular ou convenções de catalogação podem impactar negativamente no sucesso da busca entre os usuários.

Mecanismos de busca manuais ou automáticos são essenciais para o jornalismo sensível ao tempo. O processo demanda um acesso rápido, amigável e eficiente aos imensos arquivos de notícias televisivas como parte das atuais estratégias investigativas que podem criar novas histórias, mais fáceis de entender e de serem lembradas. Portanto, a fim de aprimorar as investigações jornalísticas, arquivos audiovisuais devem incluir não somente o material tradicional das transmissões de TV e as histórias fechadas mas também a cobertura ao-vivo, o material bruto e uma ampla escala de documentos descritivos. Metadados apropriados asseguram que jornalistas ou pesquisadores possam encontrar identificações mais eficientes e confiáveis de notícias relevantes em vídeo.

Outra importante ferramenta de busca digital para ser usada nos arquivos de notícias televisivas como uma fonte relevante de *Big Data* são os sistemas de reconhecimento facial. Eles consistem de uma aplicação computacional de identificação e verificação automática de uma pessoa em uma imagem digital ou em um *frame* do vídeo em um grande banco de dados audiovisual. Uma das maneiras de se encontrar pessoas não-indexadas, por exemplo, é comparando características faciais de uma imagem selecionada com banco de dados faciais. Esse recurso é tipicamente utilizado em sistemas de segurança e pode ser comparado a outras características biométricas como impressões digitais ou sistemas de reconhecimento de iris. As câmeras televisivas ou de vigilância capturam uma imagem e o sistema a converte em uma matriz de características faciais. A partir dessa matriz, a imagem original não será necessariamente observada ou identificada por uma pessoa. Assim que a câmera captura um rosto, por exemplo, ela imediatamente converte a imagem em um conjunto de algoritmos ou códigos computacionais, representando as características faciais de uma pessoa em particular. O rosto de cada um de nós é diferente, então 6 bilhões de pessoas no mundo possuem 6 bilhões de tipos de matrizes. Haveria, portanto, diferentes combinações para diferentes pessoas. O algoritmo leva somente 125 milissegundos para realizar a pesquisa nas matrizes. É possível também ter múltiplas imagens da mesma cena ao mesmo tempo²³.

Isso é uma ferramenta investigativa extraordinária, desafiadora e bastante útil para o jornalismo de dados na era digital. No passado recente, a pesquisa por rostos não-indexados em grandes bancos de dados visuais, como os arquivos de notícias televisivas, eram procedimentos custosos, trabalhosos e demorados. Além disso, jornalistas podem agora utilizar ferramentas de pesquisa audiovisuais mais específicas, tais como vocabulário controlado, *visual thesauri*, miniaturas em vídeo e *skimming* videos, que se tornaram recursos comuns em grandes bancos de dados audiovisuais.

A vantagem de um processo automático confiá-

vel é a sua capacidade de fornecer maneiras menos subjetivas de acesso à informação visual. Tais sistemas podem também ser utilizados quando o usuário quiser recuperar objetos que não foram originalmente indexados, e isso praticamente em tempo real. O processamento de imagens, o reconhecimento de imagens e a indexação em produção pode aprimorar e acelerar esse processo. A indexação com novas descrições, tais como palavras-chave, imagens e sons comprovaram ser cruciais para o jornalismo.

Apesar dos potenciais problemas e dificuldades técnicas e legais, a pesquisa por metadados audiovisuais específicos, com tecnologia multimídia em grandes arquivos audiovisuais, tende a produzir investigações de notícias ainda mais relevantes e profundas.

A tabela 1 apresenta uma descrição dos tipos de conteúdo audiovisual que estão disponíveis nos arquivos de notícias televisivas:

	Programas	Notícias	Material de estoque	Material bruto
Descrição	Boletins de notícias completos gravados tal como foram ao ar.	Notícias únicas ou componentes pré-editados de boletins de notícias.	Sequências selecionadas do material bruto e editadas em sequências reutilizáveis.	Sequências que foram gravadas para uma reportagem particular, possivelmente gravadas em fita ou no formato digital.
Estilo de catalogação	Muito sucintas ou inexistentes.	Detalhadas, incluem descrições de imagens e contexto.	Descrição detalhada de imagens e do contexto, se relevante.	Geralmente não catalogadas. Podem ser mantidas em sua integridade se estiverem relacionadas com um evento histórico maior.
Usos	Fornecer um registro histórico e exemplo de estilos de produção e de técnicas. Não é adequado para o reuso genérico.	Pode ser reutilizado em relação à história original ou em outros contextos, mas pode ser limitado devido à curta duração.	Muito versátil, pode ser utilizado em uma variedade de gêneros. Fornece cenas de maior duração e variedade do que as notícias.	Formatos longos de documentários e programas.

Fonte: <http://researcharchive.vuw.ac.nz/xmlui/bitstream/handle/10063/2235/thesis.pdf?sequence=2>

A fim de reforçar a importância e a relevância do jornalismo de dados nos arquivos de notícias televisivas é de suma importância recuperar eventos históricos recentes e paradigmáticos que confirmam a importância do uso de grandes bancos de dados como uma fonte adequada de jornalismo investigativo:

Presidente Clinton e a estagiária Monica Lewinsky



Foto 1: nessa imagem arquivada, retirada de um vídeo, Monica Lewinsky abraça o presidente Clinton no momento em que ele é saudado por simpatizantes em uma comemoração na Casa Branca em Washington, em 6 de novembro de 1996. Um relatório do advogado Robert Ray concluiu, em 2002, que promotores tinham amplas evidências de acusações criminais contra o presidente Clinton no escândalo envolvendo a então estagiária da Casa Branca Monica Lewinsky (AP Photo/APTN, arquivo).

O então chamado escândalo Lewinsky foi uma notícia sobre sexo e política que surgiu em 1998, a partir de uma suposta relação privada entre o presidente dos Estados Unidos, Bill Clinton, e uma estagiária de 22 anos da Casa Branca, Monica Lewinsky. Naquele período, uma extensiva investigação jornalística em arquivos audiovisuais de notícias usando ferramentas automatizadas de reconhecimento facial levaram à descoberta das primeiras imagens de ambos – o presidente Clinton e Monica Lewinsky – encontrando-se e conversando juntos durante um evento nos jardins da Casa Branca.

Esse foi considerado um exemplo do potencial uso de imensos arquivos de notícias televisivas juntamente com ferramentas automáticas de busca a fim de esclarecer estágios iniciais de investigações jornalísticas e judiciais. Monica Lewinsky, como era uma estagiária que trabalhava na Casa Branca, não tinha uma indexação nominal específica (seu nome não havia sido adicionado em qualquer descrição visual) em arquivos de notícias de televisão e, dessa maneira, ela não poderia ser encontrada usando tradicionais ferramentas de busca nominal. A referência cruzada visual com o “indexado” presidente Clinton demonstra o potencial dos arquivos de notícias televisivas como uma fonte de *Big Data* para futuras investigações.

²³ <http://animetrics.com/technology/frapplications.html/>

Identificação dos sequestradores do atentado de 11 de setembro



Foto 2: Os terroristas do atentado de 11 de Setembro foram identificados por câmeras de vigilância em várias cidades nos dias que sucederam os ataques. O sequestrador Hani Hanjour, que pilotou o voo 77 da American Airlines até o Pentágono, é mostrado num vídeo produzido por uma câmera de vigilância passando pela segurança no aeroporto internacional de Dulles, na manhã dos ataques. (AP Photo/Portland Police Department). Fonte: NewsComAu

Os ataques de 11 de Setembro foram uma série de atentados terroristas coordenados pelo grupo islâmico al-Qaeda contra os Estados Unidos nas cidades de Nova Iorque e na região metropolitana de Washington, em uma terça-feira, 11 de Setembro de 2001. No total, 2.996 pessoas morreram nos ataques, incluindo os 227 civis e os 19 sequestradores a bordo de quatro aviões. Foi também o incidente mais fatal envolvendo bombeiros e policiais na história dos Estados Unidos, respectivamente, com 343 e 72 mortes.

Os sequestradores Mohammed Atta e Abdulaziz al-Omari passaram pela segurança no Aeroporto Internacional de Portland, no Maine. Eles embarcaram em um voo para o Aeroporto Internacional de Logan, em Boston, onde pegaram a conexão para Los Angeles no voo 11 da American Airlines. Durante 90 minutos seguintes, 17 outros sequestradores passaram pelos controles de segurança nos aeroportos de Logan, Newark e Dulles. Todos, presumidamente, portanto facas, estiletes ou outras armas ocultas no próprio corpo ou na bagagem de mão. Embora oito tenham sido aleatoriamente selecionados para inspeção adicional e um agente no portão tenha sinalizado dois como suspeitos, nenhum deles foi barrado ao embarcar nos voos.

Em poucos minutos após os ataques, o FBI (*Federal Bureau of Investigation*) abriu a maior inves-

tigação na história dos Estados Unidos. Os suspeitos foram identificados em 72 horas. Em 27 de setembro de 2001, o FBI liberou as fotos de 19 sequestradores, junto com informações sobre suas possíveis nacionalidades e apelidos.

Bombas em Londres, 2005.



Foto 3: Os quatro terroristas capturados em CCTV na estação Luton às 07:21 da manhã no dia 07 de julho de 2005. Da esquerda para a direita: Hasib Hussain, Germaine Lindsay, Mohammad Sidique Khan e Shehzad Tanweer.

As explosões por bomba de 7 de julho de 2005 em Londres (também chamadas de 7/7) foram uma série de ataques suicidas coordenados na região central de Londres, que teve civis como alvo, usando o sistema de transporte público durante o horário de pico da manhã. Os quatro terroristas e 52 civis foram mortos e mais de 700 outros foram feridos no pior incidente terrorista do Reino Unido desde o atentado de 1988 de Lockerbie, bem como o primeiro ataque suicida ocorrido até então.

Uma imagem de câmera de vigilância mostrou três dos quatro homens supostamente responsáveis pelas explosões de 7 de julho em Londres - Mohammad Sidique Khan, Germaine Lindsay e Shehzad Tanweer - entrando na estação Luton, quando, acredita-se, faziam uma viagem de reconhecimento. Esse também é um importante exemplo de como as ferramentas automáticas de pesquisa e reconhecimento de rostos podem aprimorar o jornalismo de dados em grandes bancos de imagem de vigilância. Nesse caso em particular, o *software* trabalhou capturando um quadro de uma transmissão ao vivo para então focalizar no rosto, utilizando a distância entre os olhos como referência. Em seguida, um algoritmo utilizou as texturas e padrões únicos de cada rosto, cruzando a referência do quadro capturado, a uma taxa de milhões de comparações por segundo, com

um relevante banco de dados visual que também foi processado pelo *software*.

Esse tipo de ferramenta de pesquisa é especialmente projetada para garantir que o sistema possa trabalhar mais rapidamente e de maneira mais confiável. Quando pesquisadores orientados a dados usam sistemas e aplicações específicas, eles tendem a realizar uma tarefa com máxima eficiência e velocidade. Isso é um recurso essencial em muitas investigações jornalísticas e tais sistemas de pesquisa podem ser usados não somente em investigações de longo prazo mas também em boletins diários de notícias. Em alguns dos *softwares* mais atuais e específicos de busca automatizada, o *output* visual pode ser obtido em frações de milissegundos²⁴.

Outro recurso de uso dos arquivos de notícias televisivas para o jornalismo de dados que pode estar acessível a jornalistas durante os deslocamentos do trabalho diário é o envio de trechos de imagens em movimento para o repórter, que pode ocorrer via e-mail ou mensagens instantâneas, eventualmente disponibilizados por telefones celulares equipados com conexão de internet de banda larga, iPods ou *smartphones*.

Portanto, a internet se torna uma plataforma nova e desafiadora para a pesquisa e investigação de arquivos de notícias televisivas *online* não somente para jornalistas investigativos ou pesquisadores, mas também para repórteres ativos. Além disso, o desenvolvimento de interfaces para sistemas multimídia pode também ampliar o papel dos arquivos digitais de notícias ao endereçar questões como eficiência, credibilidade, transparência e ética no jornalismo.

12. QUESTÕES SOBRE DIREITOS AUTORAIS

Um dos problemas mais delicados sobre o uso dos arquivos de TV para jornalismo de dados está relacionado aos direitos autorais. Apesar de sua importância como fonte de informação histórica, os noticiários em muitos países, tais como os Estados Unidos, são produzidos por empresas privadas de

24 <http://animetrics.com/technology/frapplications.html/>

televisão. Ao contrário do acesso a documentos armazenados em bibliotecas públicas, por exemplo, o acesso a conteúdos em arquivos de notícias é restrito a usos comerciais.

Mas o principal desafio consiste em criar um arquivo de notícias televisivas amplo, legal e acessível. Isso requer a avaliação de potencialidades e fraquezas dos atuais e futuros modos de acesso. Questões como os modelos administrativos atuais (público *versus* privado), leis de direitos autorais, custos financeiros, arquivos não-lucrativos ou informações de “gravações em off” precisam ser considerados e desenvolvidos.

Há também a necessidade de se considerar as notícias televisivas como parte de uma “exceção” aos parâmetros de direitos autorais de acesso para pesquisas acadêmicas. De acordo com as políticas de propriedade de direitos autorais da Universidade da Califórnia, “a criação de trabalhos autorais é uma das maneiras da universidade cumprir sua missão para contribuir para o conhecimento público. A universidade encoraja a criação de trabalhos originais de autoria e a expressão livre e a troca de ideias²⁵”. Em termos de notícias televisivas, “deve haver uma exceção para permitir bibliotecas e arquivos transmitirem, eletronicamente, cópias em vídeo de noticiários por *streaming* e tecnologias similares para outras bibliotecas e arquivos elegíveis para propósitos de estudo privado, bolsas de estudo ou pesquisa sob certas condições e após um período razoável que a transmissão original tenha ocorrido.²⁶”

Sob condições legais de uso dos direitos autorais, instituições como o arquivo de notícias *Vanderbilt* e arquivos similares no mundo poderiam legalmente gravar transmissões e emprestar cópias para pesqui-

25 Texto original: “The creation of copyrighted works is one of the ways the University fulfills its mission of contributing to the body of knowledge for the public good. The University encourages the creation of original works of authorship and the free expression and exchange of ideas.” Fonte: <http://www.universityofcalifornia.edu/copyright/systemwide/pcoi.html>

26 Texto original: “there should be an exemption to allow libraries and archives to transmit view-only copies of television news programs electronically by streaming and similar technologies to other eligible libraries and archives for purposes of private study, scholarship, or research under certain conditions, and after a reasonable period has passed since the original transmission”. Fonte: idem

sas externas, mas não deveriam fornecer aos usuários permissões para fazer qualquer coisa além de ver a cópia. Isso é considerado um uso legal e, ao mesmo tempo, um tópico crucial, considerando os programas televisivos como documentos históricos. Nesse contexto, a Universidade da Califórnia de Los Angeles promove uma abordagem nova e criativa às questões de direitos autorais. Eles a chamam de “Fair Use” ou “transformabilidade” (*transformativeness*)²⁷. O propósito e a característica do uso, a natureza do trabalho autoral, a quantidade e a substancialidade do trecho utilizado e o efeito do uso no mercado ou o valor do trabalho autoral pode ser negociado entre as partes interessadas. Mas, a questão mais importante, é a garantia do acesso aberto geral para visualização de arquivos televisivos para jornalistas e pesquisadores acadêmicos.

13. REFLEXÕES CONCLUSIVAS

Praticamente cada aspecto do jornalismo atual está enfrentando uma difícil transição e, ao mesmo tempo, sendo constantemente desafiado pela crítica contundente e, algumas vezes, feroz. Exige-se que o jornalismo seja mais eficiente, relevante e transparente nos seus objetivos e métodos investigativos. Novas promessas e demandas emergem da crescente disponibilidade de fontes alternativas de investigação e técnicas como o *Big Data*, os códigos computacionais, os algoritmos e os grandes repositórios audiovisuais *online*.

Mas a pressão para produzir e difundir notícias rapidamente, de maneira barata e eficiente não deve levar ao sacrifício da busca pela verdade, excelência e independência. A checagem dos fatos foi, um dia, uma função crítica que garantia a precisão das reportagens. Ela sofreu na era dos cortes e reduções de equipes. É vital restaurar a checagem rigorosa dos fatos e, desse modo, a qualidade das notícias em vídeo e em todos os formatos. Isso pode significar o revigoramento dos processos de coleta de notícias usando *crowdsourcing* e outras abordagens criativas para ampliar o engajamento

27 http://www.crl.edu/sites/default/files/follow_up_material/2013%20RT%20News%20-%20UCLA%20Broadcast%20News%20-%20Farb%20Grappone.pdf

do público e diversificar as fontes de notícias, incluindo tanto o arquivamento formal gerenciado por instituições quanto os repositórios quase-públicos como o YouTube, que hoje abriga coleções massivas de vídeos produzidos por “repórteres-cidadãos” bem como contribuições substanciais de jornalistas profissionais e órgãos de notícias.

Este estudo tentou mostrar o potencial e estimular o jornalismo de dados destacando um dos mais importantes, mas, infelizmente, um dos mais subestimados e subutilizados de seus recursos: os arquivos de notícias televisivas. Mas, a fim de alcançar essa potencialidade, é imperativo garantir o acesso livre, transparente e direto a essas coleções, combinado-as com o desenvolvimento de novos e mais eficientes códigos computacionais, algoritmos e ferramentas digitais de pesquisa. A expansão dos direitos de acesso a esses documentos de notícias em vídeo é essencial para o uso de coleções de notícias televisivas enquanto relevante repositório de *Big Data*. Como foi indicado neste estudo, os arquivos de televisão são uma fonte principal e primária de documentos históricos que devem permitir o acesso a jornalistas, educadores, pesquisadores, estudantes e ao público em geral.

Como parte de nossa avaliação final, recomendamos algumas estratégias que devem melhorar o acesso aos arquivos de notícias televisivas para jornalistas, pesquisadores e para o público em geral. É fundamental disponibilizar instituições públicas conectadas em rede para arquivamento de bancos de dados. Para facilitar a visualização, é também importante criar possibilidades de acesso remoto a coleções digitais. Considerando a reprodução de noticiários, deve haver uma disposição para se alcançar um acordo entre as empresas de TV, a ciência, as instituições culturais privadas e públicas para aprimorar a disponibilidade de arquivos de boletins de notícias televisivas. Para o usuário final, é necessário aumentar a segurança jurídica no uso para a investigação jornalística, a pesquisa acadêmica, a educação e outros trabalhos sociais e culturais sem fins lucrativos.

Dentre as alternativas disponíveis para superar os problemas atuais sobre os arquivos de notícias televi-

sivas em várias partes do mundo, podemos considerar o uso de redes de vídeos digitais. Num futuro próximo, podemos vislumbrar que a construção de uma rede global de múltiplas fontes de arquivos de notícias em vídeo pode também se tornar um importante recurso para garantir investigações jornalísticas mais profundas em diferentes países. Novas tecnologias digitais sob o controle internacional podem garantir armazenamento em vídeo gratuito e acesso livre aos noticiários televisivos. Essas medidas podem garantir mais autenticidade, precisão e valor para o aprofundamento do jornalismo investigativo audiovisual.

Há também uma tendência em direção a um consenso social sobre a importância do arquivamento de notícias televisivas em coleções *online* de vídeos para diferentes propósitos jornalísticos, históricos e educacionais. Mas, à medida que essas novas práticas e técnicas evoluem, é essencial manter um comprometimento vigoroso com o papel central do jornalismo para a democracia. Arquivos de notícias televisivas como uma fonte de *Big Data* são ainda subexplorados e subutilizados e eles não devem ser obstruídos, ter acesso limitado ou mesmo serem “secretos”. Eles são importantes recursos para produzir histórias jornalísticas que irão engajar cidadãos a esclarecer, analisar e interpretar seu passado.

É importante garantir o livre acesso aos arquivos de notícias televisivas não somente como uma nova e importante fonte de *Big Data* para o jornalismo investigativo. Mas, esse acesso é também relevante para o cidadão engajado, a fim de que ele compreenda o mundo contemporâneo, os meandros da governança democrática e a produção de notícias pela mídia.

Essas implicações são atrativas não somente para os repórteres profissionais ou historiadores, mas para educadores do jornalismo televisivo que podem incorporar essas novas abordagens em seus treinamentos para a nova geração de jornalistas. Além disso, a colaboração próxima entre jornalistas, arquivistas de informações e dados e pesquisadores irá ajudar no avanço de uma nova geração de arquivos de notícias em vídeo com uma qualidade e acessibilidade mais intensa na era digital.

Pesquisas futuras são necessárias para determinar a extensão na qual esses desenvolvimentos caracterizam as organizações e empresas de notícias ao redor do mundo. Com o apoio de políticas públicas, bibliotecas digitais nacionais e internacionais, sob a coordenação de instituições acadêmicas como as universidades, podem patrocinar um site semelhante ao YouTube para o armazenamento e a disseminação de noticiários e outros programas relevantes de diferentes países. A internet e as novas tecnologias digitais oferecem a oportunidade de reforçar a tradição de liberdade de expressão e de informação que são vitais para o desenvolvimento social, econômico, cultural e político. Novas tecnologias podem e devem ser usadas para ampliar o direito de todos de se expressar, procurar, receber ou transmitir informações e ideias para o benefício tanto do indivíduo quanto da sociedade.

Para preservar e garantir o livre acesso aos arquivos de notícias televisivas no mundo, é necessário um grande compromisso por parte das instituições internacionais, governos, redes de televisão, jornalistas e especialmente do público em geral. Como afirma Robert Leigh, “bibliotecas de todos os tipos durante os séculos de sua existência tiveram um objetivo comum – tão normalmente aceito que raramente se torna explícito. A conservação e a organização dos recursos mundiais do pensamento registrado e o fato de os tornarem disponíveis para usuários presentes e futuros.”²⁸ (Leigh, 1950)

Não cuidar dos telejornais e demais boletins de notícias televisivas, esses documentos históricos extraordinários e comuns, é abandonar nossa obrigação como cidadãos engajados que os valoriza como uma parte significativa da herança cultural e social do mundo. O que você ou nós venhamos a fazer, será julgado pela posteridade. As dificuldades ou a total falta de acesso geral e livre aos arquivos de notícias de televisão não é somente um risco ou uma limitação para os jornalistas investigativos ou pesquisadores acadêmicos do presente. Trata-se de uma

28 Texto original: “Libraries of all kinds during the centuries of their existence have had a common objective—one so generally accepted that it is seldom made explicit. It is the conservation and organization of the world’s resources of recorded thought and fact so as to make them available for present and future users”

barreira séria para aqueles responsáveis por escrever a história do futuro.

REFERÊNCIAS

ABRAMSON, Jill. Journalism in a 24/7, Mobile, Twitter World.” Keynote address at the 2014 annual convention of the Association for Journalism and Mass Communication Education, Montreal, Canada, 2014.

AGIRREAZALDEGI, Teresa. Audiovisual documentation in the preparation of news for television news programs, *Aslib Proceedings*, Vol. 60, No. 1. pp. 47-54, 2008

AULETTA, Ken. <http://www.newyorker.com/online/blogs/currency/2014/05/why-jill-abramson-was-fired-part-three.html> Retrieved 21 May 2014.

BARONE, Antonio Fernando C. Memória. In: MARCONDES FILHO, Ciro. (ed.). *Dicionário de Comunicação* Sao Paulo: Paulus, 2009.

BRASIL, Antonio. For free access to the archives of news. Press information, n. . 695, May 22, 2012 Available in: <http://www.observatoriodaimprensa.com.br/news/view/_ed695_pelo_livre_acesso_aos_arquivos_de_telejornais>. Access: 2 July 2012.

BRASIL, Antonio; FRAZÃO, Samira. Reflexões sobre o Acesso aos Arquivos de telejornais brasileiros. *Famecos*. V. 17, n. 28. 2012.

CALDERA-SERRANO, J. Changes in the management of information in audio-visual archives following digitization: Current and future outlook. *Journal of Librarianship and Information Science*, 40(1), 13–20, 2008.

CENTRE Documentation of TV Globo (CEDOC) ‘guard’ Brazilian history. . Globo, May 22, 2012 Disponível em: <<http://redeglobo.globo.com/globo-cidadania/balanco-social-2011/noticia/2012/05/centro-de-documentacao-da-globo-cedoc-guarda-historia-brasileira.html>>. Accessed on 16 June 2012.

CODE. <http://code.org/> Retrieved 21 May 2014.

CODEACADEMY. codeacademy.com Retrieved 22 May 2014.

DERRIDA, Jacques. *Mal d’Archive*. Paris: Galillé, 1995

DERVIN, B. From the mind’s eye of the user: The sense-making qualitative-quantitative methodology. In: GLAZIER, J.D; POWELL, R. R. (org.). *Qualitative Research Information Management*. Englewood, CO: Libraries Unlimited, 1992.

DUTTON, William H. *The Fifth Estate Emerging through the Network of Networks*. Prometheus 27: 1-15, 2009.

EAGLE, Kathleen. “Digital Archiving Practices in Audiovisual Archives: An Exploration of the Use of Media Asset Management Systems in Television Broadcasting Organizations.”, 2012. <http://researcharchive.vuw.ac.nz/xmlui/handle/10063/2235> Retrieved August 18, 2014.

EDMONDSON, R. *Audiovisual Archiving: Philosophy and Principles*, 2004.

FAVIER, Jean. *Les archives*. Paris, PUF, 1959, In-16, Collection Qui Sai Je? N.805

FERRAREZI, Ludmila; ROMÃO, Lucilia Maria Sousa. File, document and memory in the discursive conception. A. *Elect. Bibliotecon. Ci Inf*, Florianópolis, n. 24, p. 152-171, 2nd sem.2007.

FOUCAULT, Michel. *L’Archéologie du savoir*, Paris, Gallimard, coll. « Bibliothèque des Sciences humaines » , 1969

FULLER, Jack. *News Values: Ideas for an Information Age*. The University of Chicago Press, 1996.

GALLUP. <http://www.gallup.com/poll/164459/trust-media-recovers-slightly-time-low.aspx> Retrieved 21 May 2014.

GREENWALD, Glenn. The Guardian. <http://www.theguardian.com/commentisfree/series/glenn-greenwald-security-liberty+world/the-nsa-files> Retrieved 21 May 2014.

HECHT, A. Verborgene Schätze? Fernseharchive und ihre Zugänglichkeit im europäischen Vergleich’(Hidden treasures”) Television archives and their accessibility in a European comparison), *Medien & Zeit* 20(3):17-21, 2005

HOOG, Emmanuel. L’INA, PUF, Paris, 2006.

L’Institut National de l’audiovisuel: Free Content and Rights Licensing the Complementary-tary Strategies. Strategic Content Alliance. Available at: <http://sca.jiscinvolve.org/wp/files/2009/07/sca_bms_casestudy_ina.pdf>. Access: 4 July 2012

LEIGH, Robert D. *The Public Library in the United States*, 1950.

JOST, François. “Écrire l’histoire culturelle avec les archives télévisuelles”, in *Archives: les enjeux de la mémoire audiovisuelle*. Les nouveaux dossiers de l’audiovisuel n. 6, Paris, INA, 2005.

KOVACH, Bill; ROSENTIEL, Tom. *The Elements of Journalism*. Crown Publishers, 2013.

KRAMP, L. Gedächtnismachine Fernsehen. Band 2: Problem unPotenziale der Fernsehherbe-Verwaltung in Deutschalnd und Nordamerika (Television, the memory machine Vol 2: Problems and potentials of television heritage management). Berlin: Akademie, 2011

KRAMP, L. Media Studies without Memory? Institutional, Economic and Legal Issues of Accessing Television Heritage in the Digital Age, retrieved from Media Practice and Everyday Agency in Europe, retrieved in Oct. 13, 2014.

LAUWERS, M. (Ed.). *Changing Roles: Media Management in the Digital Era*, 2004.

LAPOWSKI, Issie (22 May 2014). “The Startup That’s Bringing Coding to the World’s Classrooms.” *Wired*. <http://www.wired.com/2014/05/codecademy/> Retrieved 22 May 2014.

MCGUIRE, Kathleen, “History of Television Archive”, http://besser.tsoa.nyu.edu/howard/Papers/tv_archive.pdf, 2008, retrieved , 15 Oct, 2014.

MANZINI, EJ semi-structured interview: analysis objectives and roadmaps. In: International Seminar on Qualitative Research and Studies, 2, Qualitative research in debate. A-tional ... Bauru: SIPEQ 2004

MONTEIRO, Silvana Drumond; CARELLI, Ana Esmeralda. Cyberspace, memory and forgot ment. ENANCIB VIII - National Meeting on Research in Information Science, may. . Available in 2007.

MEYER, Philip. *Precision Journalism: A Reporter’s Guide to Social Science Methods*. Rowman & Littlefield, 1971.

NG, Y.; RUBIN, N.; VAN MALSSSEN, K. , Strategies for Sustainable Preservation of Born Digital Public Television. 2010.

NYTimes. <http://www.nytimes.com/2014/05/22/upshot/more-hispanics-declaring-themselves-white.html?rref=upshot> Retrieved 21 May 2014.

OLDENHAGEN, K. in Media Practice and Everyday Agency in Europe http://www.researchingcommunication.eu/book9chapters/C17_Kramp1314.pdf

PAVLIK, John. *Journalism and New Media*. Columbia University Press, 2001.

PAVLIK, John. *Media in the Digital Era*, Columbia University Press, 2008.

PEW. <http://www.pewresearch.org/fact-tank/2014/05/05/millions-of-americans-changed-their-racial-or-ethnic-identity-from-one-census-to-the-next/> Retrieved 21 May 2014.

RICOEUR, Paul, *La mémoire, l’histoire, l’oubli*, Paris, Seuil, 2000.

Society of Professional Journalists. <http://www.spj.org/shieldlaw.asp> Retrieved 22 May 2014.

TURNER, J. M. From ABC to http: The Effervescent Evolution of Indexing for Audiovisual Materials, *Cataloging & Classification Quarterly*, 48(1), 83–93, 2009.

THE GUARDIAN. <http://www.theguardian.com/uk/2011/aug/06/tottenham-riots-protesters-police>

Retrieved 21 May 2014.

UBOIS, J. 'Finding Murphy Brown: How Accessible are Historic Television Broadcasts?'

Journal of Digital Information 7(2),2005

UNESCO Recommendation concerning education for international understanding, co-operation and peace and education relating to human rights and fundamental freedoms adopted by the General Conference at its eighteenth session, Paris, 1974; http://www.unesco.org/education/nfsunesco/pdf/Peace_e.pdf Retrieved 12 Oct 2014

WIGGINS. <http://engineering.columbia.edu/ny-times-taps-prof-wiggins-chief-data-scientist> Retrieved 21 May 2014.

YAKEL, E.; KIM, Jihyun. "Adoption and Diffusion of Encoded Archival Description (EAD)," *Journal of the American Society for Information Science and Technology* (56/13), 2005.

- 52 YEE, M. M. *Moving Image Cataloging: How to Create and How to Use a Moving Image Catalog. Third millennium cataloguing.* Westport, Conn: Libraries Unlimited, 2007.

Recebido em 02 de dezembro de 2015.

Aprovado em 12 de março de 2016.