

# JORNALISMO EM AMBIENTES MULTIPLATAFORMA

Jornalismo Estruturado Por

Metadados: uma proposta de observação

**André Rosa de Oliveira.** Jornalista. Professor nos cursos de Comunicação Social das Faculdades Integradas Rio Branco. Doutor pela Universidade Metodista de São Paulo. E-mail: andrerosa.jor@gmail.com.

**Resumo simples:** De que forma a adoção de bases de dados estruturadas para o gerenciamento e distribuição de notícias produzidas para a Web potencializam a organização, recuperação e reutilização destas por outras plataformas digitais conectadas? Este trabalho relaciona esta necessidade a um conceito indispensável para qualquer nível de recuperação da informação: metadados. Argumenta que a adoção de marcações, esquemas com vocabulários controlados, ontologias formais ou outras estruturas baseadas em metadados são indispensáveis para o Jornalismo. Como suporte à discussão, apresenta o resultado de um estudo exploratório envolvendo cinco organizações de mídia, na identificação de atributos que caracterizem diferentes níveis de estruturação, resultando na proposição de um instrumento de análise baseado em funcionalidades adotadas por esses veículos.

**Palavras-chave:** Jornalismo. Internet. Metadados.

## Introdução

A evolução e reconfiguração dos meios e da prática jornalística em função da evolução tecnológica, identificada – entre outras variáveis – pela associação entre informação jornalística e o uso de softwares ou algoritmos (CORRÊA; BERTOCCHI, 2012), impulsionam a percepção da importância de entender como as tecnologias modificam ou modificarão os modos do fazer e consumir jornalismo (LIMA JUNIOR, 2011), bem como um esforço acadêmico em estabelecer linhas de investigação a partir de conceitos, entre outros, como “Jornalismo de Precisão”, sistematização seminal envolvendo ferramentas das Ciências Sociais (MEYER, 1973); “Jornalismo de Dados”, potencialização dessas práticas por meio de bases computacionais que estimulam práticas de “*Data-driven Journalism*” (GRAY; BOUNEGRU; CHAMBERS, 2012) ou, em português, “Jornalismo Guiado por

## JORNALISMO EM AMBIENTES MULTIPLATAFORMA

Dados” (TRÄSEL, 2013); e finalmente “Jornalismo Computacional” (COHEN; HAMILTON; TURNER, 2011), centrado na resolução de problemas por meio de uma visão sistêmica, movimento que produz definições como “jornalismo semântico” (GOUVÊA; LOH, 2012); “jornalismo estruturado” (CASWELL; RUSSELL; ADAIR, 2015) ou “conteúdo inteligente” (ROCKLEY; COOPER, 2012).

Nesse contexto, a Web torna-se um ecossistema fértil em narrativas digitais baseadas em experimentações e oportunidades tendo como pano de fundo a notícia apoiada em bases de dados que oferecem funcionalidades como “estocar o material produzido e preservar os arquivos (memória), assegurando o processo de recuperação das informações”, “garantir a flexibilidade combinatória e o relacionamento entre os conteúdos”, “permitir usos e concepções diferenciadas para o material de arquivo” e “armazenar anotações semânticas sobre os conteúdos inseridos” (BARBOSA; TORRES, 2013). Ao mesmo tempo, são os atributos simples que definem a Web (URLs conectadas por meio de uma rica experiência hipertextual) que explicam seu crescimento descentralizado e, de certa forma, sem regras, tornando-a um sistema aberto e complexo (BERTOCCHI, 2014).

É nessa área comum às intersecções entre os caminhos apresentados acima que reside a importância dos **metadados**, que representa o argumento principal deste trabalho e síntese de tese sobre o tema (OLIVEIRA, 2016): qualquer abordagem envolvendo o uso de dados para a produção e distribuição de conteúdos jornalísticos, incluindo suas formas de reaproveitamento especialmente por meio de algoritmos, não pode ignorar os processos de estruturação de bases de dados por meio de metadados, seja por meio de marcações ou de esquemas mais complexos, associados ao conceito de Web Semântica (BERNERS-LEE; HENDLER; LASSILA, 2001).

### **Metodologia**

Em linhas gerais, as nossas conexões na Web e os diálogos resultantes delas funcionam como documentos ligados, que seriam equivalentes a “livros construídos individualmente” (WEINBERGER, 2002). Podemos aplicar esse mesmo princípio a

## JORNALISMO EM AMBIENTES MULTIPLATAFORMA

bases de dados com conteúdo jornalístico armazenado, nas quais os metadados funcionam como um tipo de “cola” entre fragmentos. Ou, usando outra metáfora, é isso o que diferencia uma “sala cheia de livros” de uma “biblioteca”.

Em busca de uma definição ampla, podemos dizer que metadados são informações que permitem rotular, catalogar e descrever dados para serem estruturados de modo a serem compreendidos tanto por humanos quanto por máquinas. São fundamentais para a criação, descrição, organização, atualização, reutilização, validação, recuperação, preservação e recontextualização de objetos de informação. Eles representam o que pode ser descrito a respeito desses objetos em qualquer nível. Estes podem ser descritivos (voltados à descoberta e à identificação de objetos), contextuais (como as informações técnicas registradas por uma câmera digital em uma foto, bem como metadados administrativos que indicam a origem das informações e direitos de uso) ou estruturais (que definem relações entre esses objetos). Dificilmente os metadados são utilizados isoladamente: esquemas de metadados podem especificar o significado de um item, regras de armazenamento, preservação e sintaxe.

Como estruturar o conteúdo jornalístico armazenado em bases de dados por meio de metadados? Para que a notícia possa ser rotulada e categorizada de maneira a descrevê-la ou indicar a sua natureza ou propriedades (inclusive com dados que auxiliaram o jornalista a construí-la), torna-se pertinente compreender técnicas envolvendo iniciativas já propostas para relacionar elementos e gerar conexões significativas para usuários. A partir de uma abordagem interdisciplinar, levando em conta a apropriação do termo nas ciências da informação e computação, essas práticas podem ser divididas em marcações (níveis de utilização mais elementares), esquemas e ontologias (níveis mais sofisticados, que se aproximam de padrões de *linked data*) e interconexão (disponibilidade e compreensão por máquinas).

Em torno dessas práticas, foi conduzida uma investigação exploratória qualitativa em cinco veículos cujas práticas são reconhecidamente inovadoras: Globo.com, The Washington Post, The New York Times, The Guardian e BBC. Os processos que envolvem o uso de metadados nessas organizações, listados e categorizados de acordo

## JORNALISMO EM AMBIENTES MULTIPLATAFORMA

com os níveis de utilização propostos, demonstram o esforço desses grupos de mídia voltado ao enriquecimento do material jornalístico produzido – otimizando o trabalho de armazenamento, recuperação, relacionamento, distribuição de notícias. Além disso, o processo de observação contribuiu no desenvolvimento de um instrumento de análise (EISENHARDT, 1989; YIN, 2009) para futuras investigações. Uma síntese envolvendo práticas e processos está no quadro em anexo.

### **Discussão**

A multiplicidade de escolhas apresenta, num primeiro momento, níveis de complexidade – desde marcações elementares em páginas HTML até modelagem e relacionamento entre conceitos por meio de padrões semânticos. Entre esses dois pontos estão o controle na quantidade de termos descritivos, a fragmentação de células informativas, visualizações baseadas em dados armazenados em datasets externos e a disponibilização de material bruto para novas apropriações.

Paralelamente, o quadro de possibilidades técnicas sugere que funcionalidades mais sofisticadas envolvendo ontologias, triplas e padrões de *linked data* são restritas a poucas iniciativas. Essa percepção está em consonância com o ponto de vista de Everett Rogers (apud POMERANTZ, 2015), de que estas funcionalidades serão adotadas caso sejam percebidas como simples ou de utilidade evidente. Constata-se que, além do potencial de interoperabilidade das informações em ambiente digital, a combinação de escolhas se baseia, essencialmente, em propósitos claros: afinal, o que se quer com esses blocos de informação e as ferramentas para encaixá-los?

Seja qual for a abordagem envolvendo o uso de dados para a produção e distribuição de conteúdos jornalísticos, incluindo suas formas de reaproveitamento especialmente por meio de algoritmos, não pode ignorar os processos de estruturação, seja por meio de marcações ou de esquemas mais complexos. Mais do que isso, a definição de propósitos dialoga com os métodos de produção da informação jornalística, bem como com o perfil profissional na redação. Nesse aspecto, há um entendimento de que esse ambiente é menos hostil com quem se dispõe a entender como as tecnologias

# JORNALISMO EM AMBIENTES MULTIPLATAFORMA

funcionam. Cabe reconhecer ainda a interdisciplinaridade como fator primordial. Não se trata de exigir um viés tecnicista do redator que alimenta bases de dados, menos ainda de questionar o responsável por soluções de engenharia por critérios de noticiabilidade. No entanto, ao aproximá-los em equipes cujo objetivo é modelar um sistema capaz de solucionar um problema (obviamente o mesmo para jornalistas e desenvolvedores), pode haver um estímulo em cada uma das partes em explorar, contestar ideias e, conseqüentemente, desenvolver novas habilidades.

## Referências

- BARBOSA, S.; TORRES, V. O paradigma “jornalismo digital em base de dados”: modos de narrar, formatos e visualização para conteúdos. **Revista Galáxia**, São Paulo, n. 25, p. 152-164, jun. 2013.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific American**, [s. l.], p. 34-43, maio 2001.
- BERTOCCHI, D. Dos dados aos formatos: o sistema narrativo no jornalismo digital. In: ENCONTRO ANUAL DA COMPÓS, 23., 2014, Belém. **Anais...** Belém, PA: Compós, 2014.
- CASWELL, D. A.; RUSSELL, F.; ADAIR, B. Editorial aspects of reporting into structured narratives. In: COMPUTATION+JOURNALISM SYMPOSIUM, 2015, New York. **Anais...** New York: [s. n.], 2015.
- COHEN, S.; HAMILTON, J. T.; TURNER, F. Computational journalism. **Communications of the Association for Computing Machinery**, New York, v. 54, n. 10, p. 66-71, 1 out. 2011.
- CORRÊA, E. N. S.; BERTOCCHI, D. A cena cibercultural do jornalismo contemporâneo: web semântica, algoritmos, aplicativos e curadoria. **Matrizes**, São Paulo, v. 5, n. 2, p. 123-144, 2012.
- EISENHARDT, K. M. Building theories from case study research. **Academy of Management Review**, New York, v. 14, n. 4, p. 532-550, 1989.
- GOUVÊA, C.; LOH, S. Jornalismo semântico: uma visão em direção ao futuro do jornalismo online. In: LONGHI, R.; D'ANDRÉA, C. (Eds.). **Jornalismo convergente: reflexões, apropriações, experiências**. Florianópolis: Insular, 2012.
- GRAY, J.; BOUNEGRU, L.; CHAMBERS, L. **The data journalism handbook: how journalists can use data to improve the news**. [S. l.]: The Open Knowledge Foundation, European Journalism Centre, 2012.
- LIMA JUNIOR, W. T. Jornalismo computacional em função da “Era do Big Data”. **Revista Líbero**, São Paulo, v. 14, n. 28, p. 45-52, 2011.
- MEYER, P. **Precision Journalism: A Reporter's Introduction to Social Science Methods**. Bloomington: Indiana University Press, 1973
- OLIVEIRA, A. R. **Metadados como atributos da informação estruturada em bases de dados jornalísticas na web**. Tese (Doutorado em Comunicação Social). São Bernardo do Campo: Universidade Metodista de São Paulo, 2016.
- POMERANTZ, J. **Metadata**. Cambridge, MA: The MIT Press, 2015.

II SIMPÓSIO INTERNACIONAL

# JORNALISMO EM AMBIENTES MULTIPLATAFORMA

ROCKLEY, A.; COOPER, C. **Managing enterprise content**: a unified content strategy. Berkeley, CA: New Riders, 2012.

TRÄSEL, M. Jornalismo guiado por dados: relações da cultura hacker com a cultura jornalística. In: ENCONTRO ANUAL DA COMPÓS, 22., 2013, Salvador. **Anais...** Salvador: Universidade Federal da Bahia, 2013.

WEINBERGER, D. **Small Pieces Loosely Joined**. Cambridge, MA: Perseus Books, 2002.

YIN, R. K. **Case study research**: design and methods. Oaks, CA: Sage Publications, 2009.

## Quadro: Adoção de metadados em bases de dados jornalísticas

Nível de uso	Características a serem observadas	Globo.com	The New York Times	The Washington Post	The Guardian	BBC
Marcações	Utilização de marcações HTML elementares na área de identificação da página Web (" <code>&lt;head&gt;</code> "), normalmente etiquetadas por meio do atributo <code>&lt;meta&gt;</code>	Comum a todos os veículos				
	Identificação visual de campos como título, autor, data de publicação e/ou modificação e corpo do texto					
	Organização de conteúdos em nível primário por meio de editorias, categorias ou termo similar					
	Utilização de <i>tags</i> ou palavras-chave para relacionar conteúdos em nível primário ( <i>folksonomias</i> )					
Esquemas	Utilização de marcações HTML controladas na página Web, como <i>Dublin Core</i> , microformatos, microdados ou RDFa					
	Incorporação de dados externos estruturados para enriquecer suas próprias bases					
	Aproveitamento de dados externos com informações sobre conceitos (sujeitos, objetos ou lugares) para enriquecer suas próprias bases					
	Análise ( <i>parsing</i> ) e codificação de fragmentos de informação (páginas, bases de dados) com metadados por meio de <i>software</i> (codificação automática)					
	Construção de conteúdos baseados em termos controlados por meio de fragmentos de texto reutilizáveis					
Ontologias	Análise ( <i>parsing</i> ) e relacionamento de conceitos em triplas (como RDF) por meio de <i>software</i> (codificação automática)					
	Codificação de fragmentos de informação manualmente, a partir do CMS, utilizando anotações semânticas					
	Relacionamento de conceitos (sujeitos, objetos ou lugares) em ontologias por meio de triplas usando tecnologias como RDF					
Intercâmbio	Abertura de dados brutos para análise e marcação colaborativa ( <i>crowdsourcing</i> )					
	Abertura de conteúdos por meio de uma plataforma de compartilhamento					
	Abertura de ferramentas ou códigos para tratamento de dados por meio de uma plataforma de compartilhamento					
	Abertura de conteúdos por meio de API, permitindo a criação e a interoperabilidade de dados para múltiplos dispositivos e plataformas					
	Abertura de conteúdos por meio de tecnologias semânticas padronizadas pelo W3C					
	Conexão de conceitos com dados externos ( <i>datasets</i> ) por meio de tecnologias semânticas padronizadas pelo W3C					
	Desenvolvimento de agentes inteligentes capazes de reconhecer e aproveitar suas próprias bases de dados					