

## *Big data* fotográfico e o potencial de recuperação da informação Photographic Big data and information retrieval potential

Bruno Carvalho Castro SOUZA \*

**Resumo:** apresentação do conceito de *Big data* em repositórios fotográficos digitais. O texto discute o potencial de algumas técnicas para a recuperação de imagens fotográficas em repositórios digitais de grande volume.

**Palavras-chave:** *Big data*; recuperação da informação fotográfica; redes sociais; repositório digital.

**Abstract:** this is a paper about the concept of Big data in digital photographic repositories. It focuses on the potential of some techniques for the retrieving photographic images in high volume digital repositories.

**Keywords:** Big data; digital repository; photographic information retrieval; social networks.

### 1 Introdução

A discussão científica na área de acervos fotográficos vem sendo pautada pela premissa de que a fotografia carece de tratamento apropriado como documento de arquivo. Para além das disciplinas típicas do arquivo e da fotografia, a problemática vem evoluindo, passando a incorporar outras questões, por exemplo, os aspectos tecnológicos; a recuperação da informação em repositórios digitais; e a maneira como as redes sociais e seus repositórios de imagens são organizados.

Nesse cenário há pelo menos três questões que merecem ser destacadas:

- a relação entre a imagem fotográfica e o repositório digital como um todo, sob a perspectiva do valor factual do documento fotográfico *versus* o seu valor como registro em um acervo, preocupação levantada por Lacerda (2012);
- o surgimento do que Stvilia e Jørgensen (2009) consideram como coleções de leigos<sup>1</sup>, advindas da produção fotográfica de usuários de redes sociais; e
- a revolução das mídias sociais e a explosão dos repositórios digitais, bem como seus reflexos para a sociedade e para a produção fotográfica, impulsionados pela imagem de mídia social, apontadas por Hochman (2014).

---

\* Pesquisador no Centro de Pesquisa em Arquitetura da Informação (CPAI), da Universidade de Brasília. Mestre em Engenharia de Produção pela Universidade Federal de Santa Catarina (Brasil). Currículo: <http://lattes.cnpq.br/9218729989844596>; e-mail: [bruno.souza@cpai.unb.br](mailto:bruno.souza@cpai.unb.br)

<sup>1</sup> No contexto deste artigo, entende-se como leigos os indivíduos que não possuem conhecimentos formais ou experiências com os conceitos de Arquivologia, curadoria ou Biblioteconomia; ou seja, pessoas sem as habilidades e competências necessárias para realizar o tratamento técnico formal de acervos, mas que, ainda assim, mantêm, contribuem ou usufruem de repositórios digitais.

## 2 Contexto de aplicação da técnica

Lopez (2014) levanta questões pertinentes à boa parte dos acervos fotográficos no mundo, questionando como são tratadas as informações e de que forma se dá a gestão do documento fotográfico. Aponta que há uma carência de orientações claras, de parâmetros, métodos e modelos, que sejam consistentes e aderentes às normas e princípios da diplomática. Tal ausência implica a impossibilidade de uma rápida localização de imagens, em função de diferentes contextos de buscas, sem a perda do significado em relação à origem histórica (quando, onde e quem?) e contextual (a mando de quem e por que?) das fotografias.

Lopez (2000) também destaca a importância de se considerar o documento fotográfico como parte de um contexto social ao afirmar que "*o documento de arquivo – imagético ou não – está inserido na trama das relações da sociedade da qual faz parte*" (p. 24). Essa preocupação impõe a necessidade de se repensar o acervo fotográfico em função não somente de aspectos técnicos, geográficos ou editoriais, mas também do registro contextual da imagem; ou seja, a sua relação com o mundo, e com os sujeitos envolvidos em sua criação, no momento em que foi produzida. Os modelos vigentes para tratamento de acervos de documentos fotográficos não contemplam essa questão.

Sob a perspectiva tecnológica, Stvilia e Jörgensen (2009) lembram que a realidade dos repositórios digitais é recente e que, até poucos anos, acervos eram tratados quase exclusivamente por profissionais da informação, tais como bibliotecários, arquivistas e curadores, principalmente. A organização desse material geralmente é feita com base em metadados e, somente em alguns casos, a partir do contexto da instituição local que cuida do acervo. No entanto, as redes sociais promoveram todos os usuários a 'curadores' ou 'arquivistas' de sua própria produção fotográfica, o que obriga a ampliar o modelo.

Essa situação apresenta aspectos complexos em função do volume de imagens consideradas: o *Flickr*, utilizado como base empírica no estudo de Stvilia e Jörgensen (2009), possuía, em 2007, 5 milhões de usuários e mais de 250 milhões de fotografias. Segundo Smith (2015), esses números subiram para 112 milhões de usuários e 10 bilhões de imagens compartilhadas em 2015. Isso se passou em um único repositório social, que é voltado a fotógrafos, porém permite vislumbrar que a produção tem crescido exponencialmente. A *Revista Veja*, em pesquisa realizada em 2012, elaborou um levantamento sobre a quantidade de fotos produzidas no mundo e chegou à cifra de 380 bilhões (Sbarai & Honorato, 2012). Peron (2016) traz informações mais atualizadas e destaca a taxa de crescimento do repositório fotográfico do *Facebook*: em 2016, a cada segundo, aproximadamente 136 mil fotos e vídeos são enviados àquela rede social, gerando um engajamento de 3,5% — duas vezes mais do que qualquer outro tipo de conteúdo.

### 3 *Big data* fotográfico

Hochman (2014) aponta para um crescimento do que chama imagem de mídia social. Claramente se faz referência à explosão do uso das redes sociais como repositórios fotográficos, em extensões temporais e geográficas cada vez maiores:

Are these images different in kind from earlier images? What are the ways in which they offer to shape our experience of ourselves and of the world today? How do these images allow us to know the world differently, and how is this knowing different in kind from earlier kinds of understanding? (p. 1)

Esse cenário tecnológico e social demonstra o que se pode considerar como *Big data* fotográfico. Hochman ressalta a importância de se compreender que a produção das imagens em redes sociais acontece de forma contínua, em um fluxo constante. Novas imagens são incorporadas a partir de dispositivos móveis, colaborações de terceiros, compartilhamentos, releituras e uma infinidade de meios pelos quais as pessoas utilizam e reutilizam as imagens.

A existência do *Big data* fotográfico é resultado da captação massiva de documentos fotográficos. O envio automático de fotografias para repositórios digitais a partir de *smartphones*, computadores e outros *gadgets*, multiplicado pela quantidade de dispositivos e pessoas, garante massa crítica para a existência do fenômeno. Essa realidade é radicalmente diferente da estrutura comparativamente rígida e estática de qualquer acervo analógico e requer ferramentas, métodos e modelos inovadores para se compreender o ciclo da informação nesse novo paradigma, e também para fazer a gestão, organização e recuperação da informação.

Mayer-Schönberger e Cukier (2013), por exemplo, destacam a importância da mudança de percepção ao conceito de dado, pensado no sentido computacional:

Data was no longer regarded as static or stale, whose usefulness was finished once the purpose for which it was collected was achieved [...]. Rather, data became a raw material of business, a vital economic input, used to create a new form of economic value. [...] The data can reveal secrets to those with the humility, the willingness, and the tools to listen. (p. 5)

A aplicação de técnicas típicas para a gestão de *Big data* em repositórios fotográficos digitais talvez permita melhor compreender as características do repositório, suas relações, bem como determinar possíveis tendências e outros aspectos relevantes.

As principais técnicas disponíveis atualmente são:

- análise descritiva: tem o objetivo de destacar informações sobre o que acontece em determinada situação, sendo possível a geração de relatórios;
- análise diagnóstica: permite explorar os motivos pelos quais algo aconteceu em uma situação específica;
- análise preditiva: viabiliza a exploração de possibilidades futuras por meio de dados estatísticos, *data mining* e *machine learning*;

- análise prescritiva: prediz possíveis consequências para diferentes escolhas, recomendando alternativas a serem tomadas.

Essas técnicas são usualmente aplicadas a bases de dados — tanto estruturadas como não estruturadas — e é possível desenvolver métodos e modelos específicos, que permitam utilizá-las em repositórios fotográficos digitais.

#### **4 Requisitos e casos de uso**

A aplicação de técnicas de *Big data* em repositórios fotográficos pode ser viável nas seguintes situações:

- os repositórios possuem grande volume de imagens (da ordem de milhares a milhões de fotografias);
- há um fluxo constante de novos documentos fotográficos carregados para o repositório;
- é possível acessar os registros do repositório, através de API<sup>2</sup> ou outro tipo de interface de comunicação.

Os possíveis casos de uso incluem, dentre outros:

- recuperação de informações por interface de linguagem natural (recuperação semântica);
- criação de resumos automatizados sobre repositórios fotográficos digitais (indicação de palavras-chaves, temas abordados nas fotografias e outras informações potencialmente relevantes);
- análise quantitativa, descritiva e preditiva de temas e contextos derivados das informações fotográficas armazenadas no repositório.

A utilização das técnicas de *Big data* em repositórios fotográficos digitais, no entanto, ainda não dispõe de ferramentas específicas, sendo necessária a criação de modelos próprios para esse fim. As tecnologias necessárias estão disponíveis e vêm sendo alvo de forte desenvolvimento, tanto pela indústria como por pesquisadores em diversas instituições no Brasil e no mundo.

#### **5. Considerações finais**

O fenômeno do *Big data* potencializa novos conceitos, métodos e modelos para a recuperação da informação fotográfica em repositórios digitais, especialmente para aqueles caracterizados, em sua formação, por imagens de redes sociais.

A aplicação de técnicas de *Big data* em repositórios digitais, potencialmente, provoca o surgimento de relações, revelando contextos não intencionais. Possibilita a recuperação da informação fotográfica a partir da busca por aspectos semânticos, não-relacionais, capazes de extrair camadas de ordem e significado a partir de gigantescos volumes de informação acumulados nos repositórios. O armazenamento

---

<sup>2</sup> Application Programming Interface, que denota um conjunto de rotinas e padrões de programação para acesso a um aplicativo de *software*, ou plataforma, baseado na *web*.

exponencial e contínuo de imagens nos repositórios digitais é uma realidade, para a qual as técnicas de *Big data*, com mais estudos, podem permitir o desenvolvimento de soluções para a recuperação de fotografias nas redes sociais.

## Referências

- Hochman, N. (2014). The social media image. *Big Data & Society*, 1 (2). Doi: <http://dx.doi.org/10.1177/2053951714546645>
- Lacerda, A. (2012). Photographs in archives: the production and meaning of visual records. *História, Ciências, Saúde-Manguinhos*, 19 (1), 283-302. Doi: <http://dx.doi.org/10.1590/S0104-59702012000100015>
- Lopez, A. (2000). *As Razões e os sentidos: finalidades da produção documental e interpretação de conteúdos na organização arquivística de documentos imagéticos* (tese de doutorado). Recuperado de <http://eprints.rclis.org/12862/>
- Lopez, A. (2014). Políticas de acesso aos primeiros documentos fotográficos de Brasília e de sua universidade. In M. Olivera Zaldúa & A. Salvador Benítez. (Eds.). *Del artefacto mágico al píxel: estudios de fotografía* (pp. 55-69). Madrid: Fadoc/UCM. Recuperado de [http://documentacion.ucm.es/data/cont/docs/15-2014-07-24-Del%20artefacto%20ma%CC%81gico%20al%20pixel\\_standard.pdf](http://documentacion.ucm.es/data/cont/docs/15-2014-07-24-Del%20artefacto%20ma%CC%81gico%20al%20pixel_standard.pdf)
- Mayer-Schönberger, V. & Cukier, K. (2013). *Big data: a revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt.
- Peron, A. (2016). *Blog, facebook marketing: dados e estatísticas 2016* (página web). Recuperado de <http://www.allanperon.com.br/facebook-marketing/>
- Sbarai, R. & Honorato, R. (2012 abril, 15) Fotografia, o motor das redes sociais. *Veja.com* (artigo online). Recuperado de <http://veja.abril.com.br/tecnologia/fotografia-o-motor-das-redes-sociais/>
- Smith, C. (2015). 14 Interesting flickr Stats (online article). *Digital Stats Articles*. Retrieved from <http://expandedramblings.com/index.php/flickr-stats/>
- Stvilia, B. & Jörgensen, C. (2009). User-generated collection-level metadata in an online photo-sharing system. *Library & Information Science Research*, 31 (1), 54-65. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0740818808001370>