

# O *big data* como suporte às atividades forenses da perícia criminal

Gustavo Henrique M. A. da Silva

Coordenação-Geral de Tecnologia da Informação da Polícia Federal  
Setor Policial Sul, Quadra 07, Lote 23 – Brasília – DF – Brazil

henrique.ghmas@dpf.gov.br

**Abstract** *Big data have been taking a central role in the current context of information explosion, in which predominates the increasing volume, velocity and variety of information produced by people and machines. This article discusses its use in supporting activities linked to public safety and national defense, especially those related to forensic work and that can be optimized for solving crime.*

**Resumo.** *Big data vem assumindo papel central no atual contexto de explosão informacional, em que predomina o aumento do volume, da velocidade e da variedade de informações produzidas por indivíduos e máquinas. Este artigo discute sua utilização no suporte às atividades vinculadas à segurança pública e defesa nacional, especialmente aquelas relacionadas ao trabalho forense da perícia criminal e que podem ser otimizadas para o esclarecimento de crimes.*

## 1. Introdução

A atividade de investigação criminal é centralmente pautada em informações, portanto sujeita ao imperativo tecnológico e seus desdobramentos, principalmente no que se relaciona ao incremento colossal do fluxo de dados ou ao fenômeno da explosão informacional, conceituado por Saracevic (2008) como sendo o “irreprimível crescimento exponencial da informação e de seus registros”. Hoje, um dos seus principais desafios é conseguir tratar o aumento desse fluxo, decorrente de inovações tecnológicas que permitem grande capacidade de armazenamento, mas que não possibilitaram imediatos avanços em termos de processamento e análise dos dados que se apresentam em diferentes formatos e em múltiplos repositórios.

A presença corriqueira de dispositivos portáteis na rotina das pessoas, aliada ao acesso massificado à internet e à produção de informações através de múltiplos canais (redes sociais, registros de filmagem e fotográficos, bases cartográficas, georreferenciamento, *blogs*, mensagens de diversos tipos, *e-commerce*, telefonia, dentre outros) são fontes geradoras de informações que estão diretamente ligadas ao dia a dia do cidadão.

Guarino (2013) destaca, por exemplo, que na atividade forense digital os desafios de trabalhar as evidências nesse novo contexto requerem a revisão de procedimentos e de técnicas, bem como a adoção de ferramental preparado para suprir as novas necessidades, tais como Map-Reduce, árvores de decisão e redes neurais, dentre outros. Dash and Campus (2014) também cita diversos entraves em relação à

produtividade e à eficiência forense, destacando a necessidade de se adotar novas estratégias e ferramentas para alcançar o rápido processamento de informações face ao crescente aumento do poder de armazenamento ocasionado pelo fácil acesso a dispositivos computacionais por empresas e indivíduos.

O escopo deste artigo é propor, diante do cenário apresentado, um modelo informacional baseado em *big data* em que a diversidade e o volume de informações afetam diretamente os trabalhos forenses da perícia criminal, cujo objetivo finalístico é buscar a resolução dos crimes que estão sob sua responsabilidade de apuração.

## 2. A explosão informacional

A publicação *Social Media Today* (2014) indica a existência 1,15 bilhão de usuários na rede social *Facebook*, os quais adicionam 350 milhões de fotografias diariamente. Do total de usuários, 751 milhões realizam o acesso por meio de telefones móveis. Há 500 milhões de usuários registrados no *Twitter*, que produzem um movimento diário de 400 milhões de publicações. A rede social *Google Plus* possui 500 milhões de usuários. Existem quase 250 milhões de perfis no *LinkedIn*, que compartilham e registram informações a respeito de suas vidas profissionais, bem como detalham seus relacionamentos com empresas e outros tipos de organizações. Na base de dados do Instagram, estima-se a presença de 16 bilhões de fotografias, que geram mais de 1000 comentários por segundo por cerca de 130 milhões de usuários, que diariamente acrescentam mais de 5 milhões de fotografias. A estimativa é que 4,2 bilhões de pessoas acessam as redes sociais através de plataforma móveis em todo o planeta.

O site Observatório da Imprensa (2014) informou que transitaram pelas antenas de telefonia móvel do planeta cerca de 1,5 trilhão de *exabytes*<sup>1</sup> de dados por mês em média, o que representa um aumento de 81% em relação ao ano anterior.

Segundo Gantz and Reinsel (2012), entre os anos de 2005 e 2020, o universo digital vai crescer 300 vezes, partindo de 130 para 40.000 *exabytes* de dados, ou seja, haverá uma média de 5.200 *gigabytes* de dados para cada habitante da terra. Eles preveem que, somente na área de segurança, englobando as atividades de investigação criminal e inteligência militar, cerca de 45% dos dados referentes ao tema poderão ser candidatos à utilização em *big data* em 2020.

Nota-se uma tendência do crescimento de dados num ritmo bem maior ano após ano, o que exigirá habilidade, experiência e reestruturação de recursos tecnológicos para lidar com esse novo contexto. As limitações de ordem técnica devem-se à ausência de procedimentos e ferramentas capazes de entender todo o volume e a variedade de dados disponíveis e dar uma resposta em tempo hábil às demandas existentes. A busca por modelos que trabalhem com estruturas de dados mais complexas, flexíveis e que possuam melhor capacidade em se adaptar à dinâmica tecnologia atual se tornam primordiais para garantir o futuro do trabalho forense.

---

<sup>1</sup> No Sistema Internacional, 1 *exabyte* equivale a 1 milhão de *terabytes*

### 3. *Big data* e suas aplicações na atividade forense policial

Segundo Sadalage and Fowler (2013), os bancos de dados relacionais não se mostram preparados para cumprir os requisitos necessários à implementação de soluções que lidam com enormes quantidades de dados, pois apresentam pouca flexibilidade em sua estrutura. O processamento não distribuído é o conceito dominante na maioria dessas soluções, o que as tornam pouco adequadas para suportar o crescente fluxo de dados e a extração de informação exaustiva e diversificada.

Turion (2013) pontua que ferramentas *big data* podem ser úteis no apontamento, com grande margem de segurança, de situações de prevenção e solução de crimes, os quais possivelmente não ocorreriam a partir de um processo amostral, em que estaria suscetível a não descoberta de um ou outro detalhe importante ou esclarecedor.

A aplicação de novos conceitos é, portanto, essencial para a manutenção da capacidade investigatória policial, principalmente das perícias forenses. Enquanto a inteligência analítica tradicional oferece subsídios para analisar dados e transformá-los em informações no intuito de gerar vantagem competitiva para elucidar casos criminais, a utilização do *big data* possibilitaria agregar mecanismos essenciais no que se refere ao volume, velocidade, variedade, veracidade e valor dos dados [Mayer-Schönberger and Cukier, 2013].

Diante desse novo cenário, utilizar *big data* em segurança pública traz ganhos em vários processos relacionados à atividade forense, podendo ser empregadas nas seguintes tarefas [Houses of Parliament, 2014]: a) prevenção criminal: os dados de geolocalização criminal estão comumente sendo utilizados para encontrar padrões criminais relacionados à área de ocorrência, os quais podem dar suporte às equipes de investigação ao apontar os tipos criminais que ocorrem com mais regularidade em determinada área; b) predição criminal: certos tipos de crimes, como fraudes bancárias, terrorismo, roubo a bancos, dentre outros, possuem certos padrões de ocorrência. Algoritmos podem ser desenvolvidos para que tais elementos de previsibilidade sejam melhorados, permitindo criar rotinas de previsão que produzam informações importantes, buscando antever ações criminosas que coloquem a polícia em posição de vantagem em relação à ação delituosa; c) produção de informações estratégicas: a análise em *big data* pode ser utilizada para o planejamento estratégico das ações de órgãos de segurança pública. É o exemplo do projeto europeu *ePOOLICE*<sup>2</sup>, que busca relacionar dados internos da polícia com redes sociais, visando identificar novas correlações entre indivíduos que praticam crimes cibernéticos, tráfico humano e tráfico de drogas em países da União Europeia [Epoolice, 2015]; d) detecção criminal: alguns tipos de fraudes utilizam centralmente troca de dados e informações quando estão sendo cometidos. Ferramentas e softwares *big data* podem fazer com que tais processos se tornem mais eficientes e auxiliem a detectar padrões, como é o caso de crimes financeiros. Tais crimes possuem base em transações bancárias e operações financeiras diversas, o que gera grandes volumes de dados e dificulta sua descoberta; e) segurança nacional: agências de segurança nacional coletam dados de inteligência envolvendo a

---

<sup>2</sup> Early Pursuit Against Organized Crime Using Environmental Scanning, the Law and Intelligence Systems

investigação e o acompanhamento de indivíduos e organizações suspeitas de atividades danosas ao estado. Análise em *big data* fornece a capacidade de armazenar e processar grandes quantidades de comunicações eletrônicas de dados, a fim de identificar padrões ou conexões que podem indicar um comportamento ameaçador, como uma célula terrorista.

Os projetos em *big data* na área de segurança pública, embora incipientes, já são realidade em algumas instituições policiais. Segundo o *site Datameer* (2014), a cidade americana Detroit, considerada uma das mais violentas dos Estados Unidos, amenizou o problema da criminalidade ao agregar *terabytes* de dados públicos e proprietários relacionados à delinquência e seu *modus operandi*, de forma a entender o relacionamento entre indivíduos, gangues e locais de atuação.

Na Inglaterra, por meio de um sistema integrado de análise projetado para o fisco, foi possível detectar 2,6 bilhões de libras em fraudes. O sistema utiliza 28 fontes de informação que processam 1 bilhão de registros internos e externos simultaneamente (transações comerciais, impostos recolhidos, contas bancárias, dados de empresas, dentre outros), ajudando os peritos a revelar relações que não seriam descobertas por um sistema tradicional ou pelas atividades corriqueiras desempenhadas por um auditor-fiscal [Capgemini Consulting, 2013].

#### **4. Modelo Informacional do *big data* para atividades forenses**

Segundo Soares (2014), o modelo informacional do *big data* em uma organização pode ser dividido em 5 (cinco) grandes grupos de dados e insumos: dados da web e mídias sociais, *machine to machine* (M2M), grandes dados transacionais, biometria e dados gerados por indivíduos.

##### **4.1 Dados da web e mídias sociais**

As informações provenientes de redes sociais, *blogs* e outros sítios de internet estão se tornando cada vez mais importantes devido sua constante atualização e dinamismo. Diante disso, um desafio para as atividades periciais é prover integração e análises conjuntas desses dados com as bases internas da organização. No entanto, segundo Russel (2013), ainda há dificuldades de cunho técnico para o aproveitamento dessas informações. Originalmente, os dados da web e redes sociais não foram criados para o propósito de pesquisa, mas foram construídos para que sejam plataformas agradáveis de interação entre pessoas.

##### **4.2 *Machine To Machine* (M2M)**

M2M relaciona-se às tecnologias que utilizam dispositivos conectados entre si, trocando informações por meio de sensores ou medidores. Segundo Emmerson (2010), havia 50 bilhões de dispositivos frente a uma população mundial de 6,5 bilhões de pessoas à época, muitos deles interconectados através do protocolo IP e gerando informações diuturnamente.

##### **4.3 Grandes dados transacionais**

Os dados transacionais geralmente estão armazenados na organização em formatos estruturados ou não-estruturados e se relacionam diretamente a informações geradas por

sistemas tradicionais da organização. Geralmente são informações que são tratadas há anos e possuem pouco paralelismo e quase nenhum processamento distribuído.

#### 4.4 Biometria

São informações oriundas de impressões digitais, reconhecimento facial, de íris, de retina, dentre outros meios de identificação realizados por meio de características biológicas e/ou corporais.

#### 4.5 Dados gerados por indivíduos

São dados gerados por pessoas, materializados por meio de anotações, gravações de voz, fotografias, vídeos, documentos em papel, mensagens eletrônicas, e que estão apresentados, n, no formato não-estruturado. São informações valiosas, mas difíceis de serem integradas com bases de dados relacionais tradicionais.

A Tabela 1 correlaciona os grupos de dados propostos por Soares (2014) com atividades forenses que podem gerar vantagens para perícia criminal na utilização de *big data*.

**Tabela 1. Atividades forenses passíveis de utilização de *big data***

<b>Grupo de dados no <i>big data</i></b>	<b>Atividades forenses relacionadas</b>
Dados da web e mídias sociais	<ul style="list-style-type: none"> <li>• Extração de informações sobre suspeitos em sítios de busca na internet de modo a inter-relacioná-las às bases de dados existentes;</li> <li>• Pesquisa de relacionamento entre pessoas no <i>Facebook</i> ou <i>LinkedIn</i>, provando vínculo ou contato entre indivíduos ou organizações;</li> <li>• Pesquisa fotográfica na rede social <i>Instagram</i>, indicando a existência de patrimônio de um investigado (veículos, joias ou imóveis, dentre outros);</li> <li>• Localização de pessoas por meio de <i>checkins</i> realizados em redes sociais, permitindo indicar que um indivíduo está ou já esteve num determinado local, podendo provar, inclusive, encontros ou reuniões.</li> </ul>
M2M	<ul style="list-style-type: none"> <li>• Análise do padrão de deslocamento de um investigado, por meio da transmissão de sua localização em tempo real por rastreador instalado em seu veículo;</li> <li>• Interceptação telemática de uma rede <i>wi-fi</i> doméstica de um investigado e cruzamento de informações com suas atividades financeiras;</li> </ul>
Grandes dados transacionais	<ul style="list-style-type: none"> <li>• Cruzamento de informações cartorárias existentes nas bases de dados policiais que contém informações sobre investigados, indiciados, suspeitos e interrogados com dados oriundos de extratos telefônicos e bancários;</li> <li>• Cruzamento de inquéritos policiais e processos judiciais com informações constantes em laudos periciais, extraídas de sistemas de criminalística;</li> <li>• Análise e relacionamento de informações constantes em bancos de dados geridos por órgão policiais (passaporte, carteira de identidade, registro de condutores de veículos e registro de armas de fogo);</li> </ul>
Biometria	<ul style="list-style-type: none"> <li>• Cruzamento de análises de sêmen, sangue e impressões digitais com outras bases de dados;</li> <li>• Processamento de dados de reconhecimento facial extraídos de sistemas de CFTV, como catracas de edifícios e câmeras de segurança, e seu confronto com bases de dados de identificação;</li> </ul>
Dados gerados por indivíduos	<ul style="list-style-type: none"> <li>• Realização de análises em anotações, cartas, memorandos ou outros documentos coletados por meio de um mandado de busca e apreensão, procurando relacioná-las a outras bases de dados da investigação.</li> </ul>

## 5. Conclusão

O que se observa é que *big data* vem se tornando protagonista no contexto atual da explosão informacional e seu uso é essencial às atividades de segurança pública e defesa nacional, pois oferece suporte a processos de análise vinculados à prevenção e predição criminal, produção de informações estratégicas, detecção criminal e segurança nacional.

Especificamente, em relação às atividades forenses, que trabalham basicamente com processamento de dados e utilizam como insumos variadas fontes e tipos de informação, *big data* é essencial para o sucesso dos processos de trabalho conduzidos pelos peritos criminais. Viu-se que um modelo informacional permite expandir possibilidades de coleta e análise de informações das mais diversas formas e tamanhos, inclusive em tempo real, utilizando-se de dados da web e mídias sociais, M2M, dados transacionais, biometria e dados gerados por indivíduos.

Outros estudos devem ser realizados buscando soluções em temas relacionados ao dimensionamento da rede de informações, a utilização de computação em nuvem, a integração de dados estruturados e não-estruturados, formas possíveis de processamento em paralelo e formas adequadas de apresentação e visualização das informações, para que seja viável a concretização de uma plataforma forense ampla, automatizada e factível baseada em *big data*.

Além do aspecto técnico, outro ponto que merece destaque é a definição dos limites do uso das informações sigilosas manipuladas pelos governos frente à privacidade dos cidadãos. Caso implementado, um modelo analítico dessa envergadura permitirá o monitoramento de atividades diárias dos cidadãos de forma mais invasiva, por meio do cruzamento de dados de diversas fontes de informação, o que pode levar a uma sensação generalizada de vigilância estatal e a perda de credibilidade do trabalho investigativo.

## Referências

- Capgemini Consulting (2013) “An Interview with Mike Hainey. Head of Data Analytics at HMCR” [https://www.capgemini-consulting.com/resource-file-access/resource/pdf/hmrc\\_19\\_12\\_2013\\_0.pdf](https://www.capgemini-consulting.com/resource-file-access/resource/pdf/hmrc_19_12_2013_0.pdf), Agosto.
- Dash, P. and Campus, C. (2014). Fast Processing of Large (Big) Forensics Data.
- Datameer (2014) “Detroit law enforcement’s secret weapon: big data analytics. Publicado em 28 de maio de 2014.” <http://www.datameer.com/blog/customer-stories/detroit-law-enforcements-secret-weapon-big-data-analytics.html>, Agosto.
- Emmerson, B. (2010) “M2M: the Internet of 50 billion devices”. WinWin Magazine, p. 19-22.
- ePOOLICE (2015) “early Pursuit against Organized crime using environmental scanning, the Law and Intelligence systems” <https://www.epoolice.eu/EPOOLICE/about.jsp>, Julho.
- Gantz, J. and Reinsel, D. (2012) “The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east.” In: IDC iView: IDC Analyze the Future, p. 1-16.

- Guarino, A. (2013) “Digital Forensics as a Big Data Challenge”, In: ISSE 2013 Securing Electronic Business Processes, Springer Fachmedien Wiesbaden, p. 197-203.
- Mayer-Schönberger, V. and Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think, Houghton Mifflin Harcourt.
- Observatório da Imprensa (2014) “Mundo consome 1,5 exabyte por mês em dados móveis.” [http://www.observatoriodaimprensa.com.br/news/view/\\_ed785\\_mundo\\_consome\\_15\\_exabyte\\_por\\_mes\\_em\\_dados\\_moveis](http://www.observatoriodaimprensa.com.br/news/view/_ed785_mundo_consome_15_exabyte_por_mes_em_dados_moveis)>, Julho.
- Russel, M. A. (2013) Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More, O'Reilly Media, Inc.
- Sadalage, J. P. and Fowler, M. (2013), NoSQL Essencial: Um Guia Conciso para o Mundo Emergente da Persistência Poliglota, Novatec, São Paulo.
- Saracevic, T. (1996) “Ciência da informação: origem, evolução e relações”, In: Perspectivas em ciência da informação, vol. 1, nº 1, p. 41-62.
- Soares, S. (2012), Big data Governance: An Emerging Imperative. MC Press, LLC.
- Social Media Today (2014) “Top 5 Social Media Predictions for 2015”, <http://www.socialmediatoday.com/content/top-5-social-media-predictions-2015>, Agosto.
- Turion, C. (2013), Big data, Brasport, Rio de Janeiro.