



BIG DATA, DATA SCIENCE E SEUS CONTRIBUTOS PARA O AVANÇO NO USO DA OPEN SOURCE INTELLIGENCE

Danielle Sandler dos Passos¹

1 NOVA Information Management School, Instituto Universitário de Lisboa (ISCTE), Universidade Nova de Lisboa

RESUMO

Diante do crescente avanço tecnológico e da ampla disponibilização de dados pelos meios de comunicação aberta, vê-se surgir a necessidade de novos mecanismos que auxiliem a captar e analisar corretamente tais informações, tirando o máximo de proveito que a *Open Source Intelligence (OSINT)* pode oferecer. O objetivo de estudo deste artigo é descobrir as vantagens do uso da *OSINT* pelas agências de inteligência, organizações e empresas, e como os mecanismos de Big Data e Data Science podem auxiliar na sua propagação e torná-la cada vez mais eficaz.

Palavras-chave: Inteligência; Open Source Intelligence (OSINT); Big Data; Data Science.

1. INTRODUÇÃO

Com o surgimento e a ampla difusão da internet, os veículos de comunicação tomaram forma e dimensão novas. Hoje, tem-se uma enorme gama de informações à disposição, surgindo o que é chamado de *Open Source Intelligence (OSINT)*, que se refere à Inteligência, no sentido de informações, obtida por meio de dados disponíveis para o público em geral, por meio de veículos de comunicação tais como jornais, sites, blogs, revistas científicas, TV, entre outros.

Entretanto, de nada basta ter milhares de informações à disposição se ninguém souber como fazer bom uso delas. Em prol dessa necessidade de valorização e tratamento das informações, atualmente existem duas importantes ferramentas: o *Big Data* e a *Data Science*.

O *Big Data* pode, modestamente, ser definido como um grande e complexo conjunto de dados, cujos métodos de processamento tradicionais seriam insuficientes para seu tratamento – que inclui processos como análise, captura, pesquisa, compartilhamento, armazenamento, transferência, visualização e segurança das informações.

Já a *Data Science* é descrita como a ciência responsável pela análise e utilização de dados que incorpora técnicas e teorias de diversas áreas, como lógica, matemática, estatística, computação, engenharia e economia.

Assim, diante do exposto, este artigo pretende mostrar os benefícios da incorporação da *Open Source Intelligence* no dia a dia e como o *Big Data* e a *Data Science* podem auxiliar nesse processo, tornando a *OSINT* cada vez mais útil e eficaz nas tomadas de decisão.

2. O QUE É INTELIGÊNCIA E O QUE A DIFERE DE DADOS, INFORMAÇÃO E CONHECIMENTO

Dentre as muitas definições de inteligência, o primeiro conceito (surgido nos campos corporativo e de espionagem) a descrevia como o produto da coleta, análise, avaliação, integração e interpretação de todas as informações disponíveis, podendo afetar ou não a sobrevivência e o sucesso da organização (Eells et Nehemkis, 1984). Já numa definição mais ampla, em que a inteligência é similar a conhecimento e informação, é descrita como toda informação coletada, organizada e/ou analisada para atender à demanda de um tomador de decisão (Cepik, 2002). E, numa definição mais restrita, é a coleta de informações sem consentimento, cooperação ou mesmo conhecimento por parte dos alvos da ação (Cepik, 2002).

Contudo, é importante perceber que, independentemente da definição escolhida, todas as atividades ligadas à inteligência visam à produção de conhecimento baseado em dados devidamente selecionados, avaliados, interpretados e, por fim, expostos de maneira útil para o processo decisório.



Diferentemente do que muitos podem pensar, inteligência não é o mesmo que conhecimento ou informação. Informação é o dado contextualizado; conhecimento é o resultado da análise da informação com base na aprendizagem e na experiência do indivíduo; e inteligência é a utilização prática do conhecimento, quando acionado numa tomada de decisão. Enfim, toda inteligência é informação, mas nem toda informação é inteligência (Lowenthal, 2012).

Além disso, o processo da criação de inteligência ainda pode ser categorizado conforme a fonte de coleta de dados — OSINT (inteligência de fontes abertas), HUMINT (inteligência de fontes humana), SIGINT (inteligência de sinais) e IMINT (inteligência de imagens) — e, independentemente da fonte escolhida, após a coleta, se faz a verificação, análise e tratamento dos dados, para torná-los úteis no processo de tomada de decisão¹.

Em síntese, o processo é sempre o mesmo: coleta de informação conforme necessidade e análise, e reporte para o tomador de decisão. Entretanto, ao longo dos anos, viu-se que as fontes abertas passaram a ser de grande importância para o processo. Com o surgimento da internet e o avanço tecnológico, o mundo passou a ter acesso e a compartilhar milhares de informações em tempo real — com isso, deu para perceber quão errônea é a ideia de que só informações sigilosas é que são valiosas. Passa-se, então, a dar destaque a *Open Source Intelligence*.

3. OPEN SOURCE INTELLIGENCE (OSINT)

É um conceito amplo e em franca expansão no ambiente das agências de inteligência e entre empresas e órgãos governamentais. Ele remete à ideia de utilização de fontes abertas para recolhimento de informações. A OSINT é definida como a análise baseada na “obtenção legal de documentos oficiais sem restrição de segurança, da observação direta e não clandestina dos aspectos políticos, militares e econômicos da vida interna de outros países ou alvos, do monitoramento da mídia, da aquisição legal de livros e revistas especializadas de caráter técnico-científico, enfim, de

um leque mais ou menos amplo de fontes disponíveis cujo acesso é permitido sem restrições especiais de segurança” (Cepik, 2003).

Entenda-se aqui como fonte aberta, veículos de informação como mídias (jornais, revistas, rádio, TV), dados públicos (relatórios governamentais, orçamentos públicos) e produções profissionais e acadêmicas (artigos, *papers*, simpósios, conferências). E ainda: Literatura cinzenta (pesquisas científicas e tecnológicas e outros materiais com distribuição limitada), observação de terceiros e conteúdo na Web (qualquer um se torna fonte de informação) e informações geoespaciais (fotos de satélites, mapeamento de campo) (Brito, 2006).

Na OSINT, os processos envolvidos visam à coleta de informações de fontes abertas e seu tratamento. Por fim, seu resultado será o produto de um raciocínio baseado e contextualizado a um fato ou ação.

Como pioneiro na utilização da OSINT, pode-se citar a *Foreign Broadcast Information Service* (FBIS), organização norte-americana situada na Universidade de Princeton que, durante a II Guerra Mundial, recolhia informações em noticiários internacionais como fonte de inteligência e, já na Guerra Fria, monitorizava publicações oficiais da União das Repúblicas Socialistas Soviéticas (URSS). Após o fim da Guerra Fria, o FBIS perdeu um pouco sua função, pois teoricamente não existia uma ameaça ou inimigo real aos EUA. Entretanto, com os atentados de 11 de setembro (2001)², a utilização das fontes abertas voltou a ter destaque. Afinal, após o ocorrido, viu-se que muitas informações que poderiam ter ajudado a prever (e, quem sabe, até a evitar) os atentados estavam à disposição do grande público.

Desde 2001, a NATO³ defende a utilização ampla da Open Source Intelligence e, nesse âmbito, criou os conceitos Open Source Data (OSD) e Open Source Information (OSI). Ambos se aludem à informação antes de sua análise, logo que é capturada. O OSD é utilizado para designar elementos, como fotografias e imagens de satélite comerciais, e o OSI se refere às informações provenientes de meios de comunicação social, relatórios, livros e demais publicações do gênero. Para a organização, OSINT é “a informação que foi deliberadamente descoberta, discriminada, destilada e disseminada

1 Processo conhecido como Ciclo da Inteligência, descrito por Johnson, R. em *Analytic Culture in the US Intelligence Community – an Ethnographic Study*, 2005. Consiste nas seguintes etapas: 1) Planejamento e direção: gestão de todo o esforço do processo e a determinação dos requisitos de escolha dos dados, baseados na demanda apresentada; 2) Coleta: captura de dados brutos (ainda não analisados e tratados), conforme demanda; 3) Processamento: análise e tratamento dos dados brutos, de forma que possam ser utilizados na tomada de decisão; 4) Análise e produção: verificação da confiabilidade, validade e relevância, das informações recolhidas; 5) Disseminação: compartilhamento do conhecimento produzido com o público-alvo.

2 Série de ataques terroristas contra os Estados Unidos, coordenados pelo grupo terrorista islâmico Al-Qaeda, que resultou na colisão de 2 aviões contra as Torres Gêmeas (edifícios do complexo empresarial do World Trade Center, em Nova Iorque), levando a morte de centena de pessoas.

3 North Atlantic Treaty Organization, também conhecida por OTAN (Organisation du Traité de l'Atlantique Nord), criada em 1949 com o objetivo de garantir a defesa coletiva dos países membros (atualmente 28) em resposta a ataques sofridos.



por uma audiência selecionada, de modo a responder a uma questão específica”.

Se a inteligência provém de informações descobertas, discriminadas, destiladas e difundidas para o tomador de decisão (Steele, 2006) (independentemente de serem de fonte aberta ou sigilosa) e se são aceitas tanto a definição de inteligência como conhecimento ou informação analisada quanto a de um segredo ou informação secreta, por que então os serviços secretos e seus espões estão perdendo espaço para OSINT?

Como resposta, podem-se elencar alguns motivos, dentre eles destacam-se a constatação de que há milhares de informações relevantes nas fontes abertas — como o fato de que, de 1998 a 2008, passou-se de 15 *links* terroristas presentes em *websites* para mais de 4500 — e a considerável diminuição de custos para obtenção das informações, afinal elas estão à disposição de quem as quiser.

Todavia, as informações coletadas de nada servem se não forem corretamente filtradas, analisadas e validadas. Assim, tornam-se de suma importância os processos utilizados e os analistas envolvidos no processamento dos dados. A OSINT só será benéfica para o processo se houver a correta implementação e investimentos em sistemas, estrutura e tecnologia, com agentes qualificados, devidamente treinados para descobrir fontes adequadas de informações, definir quais dados são relevantes para suprir a demanda requerida e analisá-los. Afinal, hoje o grande problema não é falta de dados, mas sim a correta análise deles.

Diante disso, surgem softwares, tecnologias, conceito e culturas envolvidos no processo de inteligência. Dentre esses, os que vêm ganhando grande destaque são o *Big Data* e a *Data Science*.

4. BIG DATA

O termo *Big Data* surge no começo dos anos 90, na NASA⁴, com o intuito de descrever o conceito de conjuntos de dados grandes e complexos, onde os sistemas e estruturas computacionais utilizados até então não seriam suficientes para a correta captura, processamento, análise e armazenamento informacional. Assim, pode-se descrevê-lo como a utilização de sistemas e tecnologias eficientes na valorização de grandes conjuntos de dados, tornando-os mais precisos e auxiliando na mitigação dos riscos envolvidos no processo de tomada de decisão.

O *Big Data* conta com diversas tecnologias e algoritmos que são implementados a grandes bancos de dados (estruturados ou não) com intuito de efetuar correta captura, análise, processamento e disseminação das informações, conforme demanda e objetivo traçado ao início do processo e frequentemente reanalisado. Seu principal objetivo é tornar as informações úteis para o processo decisório.

Um estudo realizado pela OBS (*Online Business School*) mostrou que, de 2004 a 2014, gerou-se mais dados do que em todo período anterior na história. Isso confirma o que disse Peter Norvig, diretor de pesquisas do Google: “Nós não temos melhores algoritmos. Temos apenas mais dados”.

Nessa conjuntura, percebe-se o papel primordial da informação. Entretanto apesar do que Norvig pode dar a entender, só tê-la não basta. O receptor precisa interpretá-la ao seu favor. E é nessa conjuntura que o *Big Data* se torna tão importante. Por meio dos preceitos de volume, variedade, velocidade, veracidade e valor⁵, o *Big Data* é capaz de armazenar um elevado número de informações diversas, efetuar rapidamente sua análise e ainda checar sua veracidade, o que o permite agregar valor ao processo de tomada de decisão, tornando-o mais eficaz e eficiente.

Associado à OSINT, o *Big Data* vem conseguindo mapear padrões de comportamento e tendências. O projeto *Google Flu Trends* é um bom exemplo disso. Por meio dele, conseguiu-se identificar uma epidemia de gripe usando como fonte de informação os dados que os usuários jogavam no sistema de busca do *Google*. Mapeando as áreas geográficas onde se buscou na web palavras referentes à gripe, ponderou-se que uma epidemia de gripe estava realmente acontecendo nessas regiões. O *Big Data* também vem auxiliando na identificação de padrões de comportamentos terroristas presentes nas mídias sociais, o que é de grande valia em sua prevenção e ataque.

Além disso, pode-se observar que, com o *Big Data*, vieram mudanças significativas na maneira a análise de dados é pensada e executada. A primeira mudança é que ao tratar grandes volumes de dados, a percepção de como olhar os dados muda. Ou seja, mudando a escala, nossa percepção também se altera. A segunda mudança é que, por se tratar de enormes quantidades de informações, o N da amostra é gigantesco, assim deixa-se de perseguir a exatidão e passa-se a se preocupar com a tendência. Isso porque, ao se trabalhar com números pequenos busca-se uma exatidão dos dados, pois só assim existe a certeza de um resultado. Porém, com amostras muito grandes, a tendência já direciona para o resultado. E, por fim, a terceira mudança é a libertação dos efeitos de causalidade para a dedicação às correla-

4 National Aeronautics and Space Administration, agência do governo americano responsável pela pesquisa e desenvolvimento de tecnologias e programas espaciais.

5 Conceito dos 5 Vs criado, em 2001, pelo analista de informação Doug Laney, para descrever o *Big Data*.



ções. Afinal, num volume grande de dados as causalidades serão em pequeníssimas quantidades e, se não o forem, serão consideradas correlações que indicarão uma tendência.

Assim, pode-se ver que as associações e análises presentes ao longo dos processos envolvidos no *Big Data* seriam impossíveis de se executar se não fossem as tecnologias e os sistemas atuais, sendo os métodos estatísticos por si só ineficientes para tal. Entretanto, o *Big Data* só será bem-sucedido, sem déficits analíticos causados por desinformação ou pela má qualidade do dado, se os parâmetros e objetivos forem bem estabelecidos e se houver analistas preparados e especializados em suas áreas temáticas. É aqui que entra a Data Science.

5. DATA SCIENCE

Pode ser definida como um conjunto de técnicas utilizadas no processamento e análise de dados, com intuito de fornecer informações para decisões inteligentes. Para tanto, mescla-se diversas áreas do conhecimento, desde conceitos simples de estatística até complexos algoritmos.

Seus analistas são conhecidos como *data scientists* e é desejável que tenham qualificações na área de tecnologia da informação (TI) para capturar eficientemente e em tempo hábil os dados; conhecimentos matemáticos e estatísticos para definir os modelos e algoritmos a serem utilizados e entender suas implicações e resultados; e, por fim, conhecimento do negócio para poder traduzir os resultados em informações que auxiliem o tomador de decisão.

O processo de *Data Science* é bem semelhante ao de *Big Data* — inicia com a coleta dos dados por meio do correto questionamento do problema/objetivo. Segue com a análise dos dados, com a visualização e aplicação de técnicas e algoritmos, e finaliza com a comunicação dos resultados.

Entretanto, ao longo do processo, surgirá a necessidade de novos dados — alguns serão descartados e erros de análise aparecerão. É por isso que os analistas necessitam ter um amplo *know how* em diversos campos e ciências, pois devem fazer as perguntas certas, capturar os dados certos e ter a correta percepção de como proceder ao longo do processo para, ao final, dados se transformarem em inteligência.

6. CONCLUSÃO

Com a ampla disposição de dados gerada pela “democratização da informação” e pelo avanço tecnológico e sua popularização, a *Open Source Intelligence* despontou entre as fontes utilizadas para obtenção de dados.

Dentre as vantagens da sua utilização, destaca-se o alto grau de oportunidade, com muitas informações à disposição e o baixo custo para obtê-las. Em época de crises e ajustes orçamentários, seu uso se torna mais atraente, possibilitando a ampliação das possibilidades dos serviços de inteligência. Entretanto, a quantidade exagerada de dados, a qualidade duvidosa das informações e a falta de confiança nas fontes podem desabilitar os benefícios do uso das fontes abertas. É aí que surge a importância de conciliar as práticas de *Big Data* e da *Data Science* ao uso da OSINT.

A utilização do *Big Data* e da *Data Science* em processos de inteligência atribui-lhe uma mais valia, pois permite ganhos de eficiência relacionados a custo, inovação e produtividade. Isto porque, para execução dos seus processos, designam analistas com amplo conhecimento e experiência em diversos campos, além de contar com avançadas tecnologias, sistemas e estruturas que permitem a captura e manipulação das informações necessárias a sua demanda, transformando-as em inteligência — informação útil ao processo de tomada de decisão.

Enfim, fica evidente que os serviços de inteligência, empresas e organizações se beneficiam muito com o uso do *Big Data* e da *Data Science* na manipulação das informações provenientes de fontes abertas. Eles tornam a *Open Source Intelligence* uma fonte ampla, segura, barata e eficiente de informações, o que só contribui para o resultado das suas atividades e provém vantagem competitiva nos processos de tomada de decisão.

REFERÊNCIAS

- Afonso, L. (2006), “Fontes abertas e Inteligência de Estado”, Revista Brasileira de Inteligência, No. 2, disponível em: www.abin.gov.br/modules/mastop_publish/?tac=Fontes_abertas_e_Inteligencia_de_Estado (Acesso em 01 de junho de 2015).
- Best, C. (2008), “Open Source Intelligence”. Joint Research Centre, disponível em: media.eurekalert.org/aaasnewsroom/2008/FIL_00000000010/071119_MMDSS-chapter_CB.pdf (Acesso em 05 de junho de 2015).
- Brito, V. (2006), O Papel Informacional dos Serviços Secretos, Dissertação de Mestrado em Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, MG.
- Cepik, M. (2003), Espionagem e democracia, 1 ed., FGV, Rio de Janeiro, RJ.
- Cepik, M. Inteligência e Políticas Públicas: dinâmicas operacionais e condições de legitimação, Security and Defense Studies Review, Nº 2, vol. 2. Rio de Janeiro, 2002.



- Ghiggi, L. et SEBBEN, S. (2009), "Inteligência", Dossiê Temático N°06, disponível em www.ufrgs.br/nerint/folder/artigos/artigo76.pdf (Acesso em 02 de junho de 2015).
- Eells, R. et NEHEMKIS, P. (1984). *Corporate intelligence and espionage: A blueprint for executive decision making*, 1 ed., Macmillan, New York, NY.
- Gonçalves, J. (2013), *Atividade de Inteligência e Legislação Correlata*, 3 ed., Impetus, Niterói, RJ.
- Johnston, R. (2005), "Analytic Culture in the US Intelligence Community – an Ethnographic Study", disponível em: www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/analytic-culture-in-the-u-s-intelligence-community/analytic_culture_report.pdf (Acesso em 25 de maio de 2015).
- Leite, S. (2014), "O Emprego das Fontes abertas no Âmbito da Atividade de Inteligência Policial", *Revista Brasileira de Ciências Sociais*, Vol. 1, No. 5, Brasília, DF.
- Lowenthal, M. (2012), *Intelligence: From Secrets to Policy*, 5 ed., CQPress, Washington, DC.
- Mendes, G., Moresi, E., Silva, W. (2010), "Estudo sobre Portais Públicos como Fontes Confiáveis para Inteligência de Fontes Abertas", artigo apresentado no COVIBRA 2010: Congresso Virtual Brasileiro – Administração, 19 - 21 de novembro, 2010, disponível em: www.convibra.org/2010.asp?ev=71&p=&lang=en (Acesso em 20 de maio de 2015).
- Mendes, G. et MORESI, E. (2012), "Operações de Informação: um estudo sobre o desenvolvimento de doutrina aplicada à prevenção à fraude", *Sistemas, Cibernética E Informática*, Vol. 9, No. 1, Brasília, DF.
- North Atlantic Treaty Organization. (2001), "Open Source Handbook", Vol. 1, disponível em: www.oss.net/dynamaster/file_archive/030201/ca5fb66734f540fbb4f8f6_ef759b258c/NATO%20OSINT%20Handbook%20v1.2%20-%20Jan%202002.pdf (Acesso em 20 de maio de 2015).
- Steele, R. (2006), "Open Source Intelligence", *Forbes*, disponível em: http://www.forbes.com/2006/04/15/open-source-intelligence_cx_rs_06slate_0418steele.html (Acesso em 19 de maio de 2015).