lavits ****** 2017
5to Simposio Internacional LAVITS
Vigilancia, Democracia y Privacidad
en América Latina: Vulnerabilidades y resistencias

ISSN 2175-9596

A SOMBRA DO DARK DATA E OS MERCADOS DE VIGILÂNCIA

La sombra del Dark Data y los mercados de vigilancia

The Dark Data Shadow and the surveillance markets

André Ramiro[a]
Marcílio Braz[b]

[(a)] E-mail: andrebramiro@gmail.com.
[(b)] E-mail: marciliobrazjr@mail.com.

Resumo

O capitalismo de vigilância, teoria cunhada por Shoshana Zuboff, vem assumindo o protagonismo e ditando as (co)operações econômico-políticas entre o setor privado e os Estados. Em um desdobramento do capitalismo informacional de Manuel Castells, uma nova economia baseada em inovadores modelos de negócio surgidos com a crescente utilização de Big Data vêm gerando repercussões que atingem cada vez mais questões como a garantia de direitos fundamentais, em especial à privacidade, sobretudo diante das cada vez mais sofisticadas técnicas de coleta, utilização massiva de dados e de-anonimização. Nesse cenário, é seguro que há de se rediscutir (ou ao menos problematizar) o conceito de privacidade da forma como, atualmente, é conhecido. Tal como as *dark energy e dark matter*, resultantes do Big Bang, as quais somadas compõem quase a totalidade de toda a matéria/energia existentes no Universo e que detém a fundamental função de evitar a dispersão total da matéria, a *dark data* vem ganhando força como parte igualmente majoritária de dados gerados, coletados e armazenados por empresas cujo modelo de negócio se baseia em exploração de Big Data e em vigilância permanente. *Dark data* pode ser considerada como os dados não estruturados, que não são explorados ou aproveitados, inúteis - até então - do ponto de vista econômico. Diante de novos avanços tecnológicos, um dos possíveis cenários, como o de de-anonimização de dados pessoais, a questão irá gerar repercussões quanto à violação de privacidade, apesar da figura da pseudoanonimização. Este artigo pretende, a partir de uma conceituação de *dark data*, abordar seu presente uso comercial, levando em consideração as possibilidades de de-anonimização, e relacionar o fenômeno às diferentes abordagens em relação a proteção de dados pessoais nos Estados Unidos, na Europa e, finalmente, situar a previsão atual brasileira, encontrando norte no Marco Civil da Internet e, em especial, no Projeto de Lei 5276/2016, relativo à Proteção de Dados Pessoais.

Palavras-chave: Dados pessoais; Dark Data; Data-driven market; Surveillance capitalism; Big data.

Resumen

El capitalismo de vigilancia, teoría desarrollada por Shoshana Zuboff, tiene asumido el papel de protagonista dictando las reglas económico-políticas de cooperación entre el sector privado y los gobiernos. Desde el capitalismo informacional de Manuel Castells, una nueva economía basada en modelos de negocio innovadores, desarrollados desde el uso creciente de Big Data, viene afectando cada vez más las cuestiones entorno la garantía de derechos fundamentales y la privacidad, sobretodo ante las, aún más sofisticadas, técnicas de recogida de datos y de-anonimización. Ante este escenario, se ha vuelto imprescindible rediscutir (o a lo mejor problematizar) el concepto de privacidad tal cual es conocido hoy. Tal cual el dark energy y el dark matter, resultantes del Big Bang, que componen casi la totalidad de materia/energía existente y evitan la dispersión total de la materia, el dark data viene ganando fuerza y asumiendo la parte más grande de los datos recogidos y almacenados por empresas cuyo modelo de negocio tiene por base la exploración de Big Data y la vigilancia permanente. El dark data puede ser entendido como los datos no estructurados, que no son explorados o aprovechados – inútiles hasta ahora – desde la perspectiva económica. Frente a los nuevos avanzos tecnológicos, uno de los posibles contextos, como de la de-anonimización de datos personales, la cuestión provocará repercusiones no que toca la violación de la privacidad, a pesar de la figura de la pseudo-anonimización. Este artículo desea, partiendo de una concepción de dark data, tratar acerca de su presente uso comercial, manteniendo en cuenta las posibilidades de de-anonimización y establecer relación con la protección de datos personales en el EE.UU., en la Europa y, por fin, situar el pronóstico de la situación brasileña, trazando un norte en el Marco Civil da Internet, y en especial en el proyecto de ley 5276/2016, referente a la protección de los datos personales.
Palabras clave: *Datos personales; Dark Data; Data-driven market; Surveillance capitalismo; Big Data.*

Abstract

The surveillance capitalism, theory developed by Shoshana Zuboff, has been assuming the leading role and dictating the economic and political (co)operations between the private sector and the states in general. In a deployment of Manuel Castells' "informational capitalism", a new economy based on innovative business models stemming from increasing use of Big Data has been generating repercussions that are rising issues regarding the guarantee of fundamental rights, in particular privacy, especially in the face of more and more sophisticated techniques used for collecting and massive data usage and its deanonymisation. In this scenario, this concept of privacy necessarily has to be rediscussed (or at least, problematized) in the way that it is currently known. Like dark energy and dark matter, resulting from the Big Bang, that together compose almost all matter/energy existent in the Universe, and which has the fundamental function avoiding the total dispersion of matter, the dark data has been gaining strength as part of the majority of data generated, collected and stored by companies whose business model is based on Big Data exploitation and permanent surveillance. Dark Data may be considered as unstructured data, which are not exploited or used, disposable - until then - from the economic point of view. In face of technological advances, one of the possible scenarios, such as deanonymisation of personal data, will still generate repercussions regarding the violation of privacy, despite of the figure of pseudo-anonymization. This article aims, starting from a dark data conceptualization, to address its present commercial use, taking into account the

possibilities of deanonymisation, and to relate the phenomenon to different approaches with regard to the personal data protection in the United States, Europe and, finally, to locate the Brazilian current legal provision, finding north in the Brazilian Civil Framework of the Internet and, particularly, in the Law Project 5276/2016, on the Protection of Personal Data.
Keywords: *Personal data; Dark Data; Data-driven market; Surveillance capitalism; Big Data.*

CONCEITO E DESAFIOS

Partindo do princípio do "capitalismo informacional" elaborado por Manuel Castells (2016) – em que o mercado da era da informação, globalizado, rearranja os processos de produtividade e produção, inaugurando novos mercados e matérias-primas, aprofundando a competitividade – é possível traçar uma linha crítica até Shoshana Zuboff (2015) e a teoria do "capitalismo de vigilância", focada precisamente no mercado de dados gerado pelo uso massivo das mídias digitais.

Neste cenário, um novo "*data-driven market*" vem se sofisticando, favorecido pela cada vez mais avançada ciência de dados e cada vez maiores níveis de uso de Big Data, fenômeno que não se apresenta livre e com sérios questionamentos sobre ética e transparência (Boyd & Crawford, 2015). Este cenário se baseia, em certa medida, em explorar vulnerabilidades de privacidade de usuários, em um câmbio lucrativo de dados pessoais em troca de serviços "gratuitos". Portanto, a constante vigilância implica no crescimento dos bancos de dados que se tornam cada vez mais valiosos, estimulando mais vigilância, caracterizando o âmago de modelos de negócio como os da Amazon, Apple, Microsoft, e especialmente Facebook e Google. Nesses últimos, a renda advinda da publicidade digital chega a, respectivamente, incríveis 97% e 88% de seus faturamentos anuais, totalizando 106 bilhões de dólares ou 57,6% do total do mercado em publicidade digital no mundo, este impulsionado em grande parte pelo comércio de dados (Desjardins, 2017a; Hernandez, 2016).

O sucesso da mercantilização de dados pessoais explica em 2016 as 5 empresas mencionadas serem as mais valorizadas no mundo (Desjardins, 2017b). Em 2011, apenas a Apple fazia parte deste seleto grupo. Um banco e três companhias de petróleo foram desbancados: "os dados são o novo petróleo" (The Economist, 2017).

Por mais avançado que estejam os processos de mineração e tratamento de dados no estágio atual, pode-se dizer que este potencial ainda esteja em processo de desenvolvimento, visto que na efetiva utilização advinda da coleta de dados, apenas 10% dos dados não estruturados são analisados,

restando uma percentagem esmagadora de dados a serem processados (Pal, 2015).

Podemos dizer, portanto, que há uma gigantesca monta dos dados, que não se figura à luz do tratamento e comercialização de dados, mas é, por motivos mais adiante expostos, coletada. Esta é mais conhecida, no mercado e na ciência de dados, como ***dark data*** e oferece um potencial inominável de aprofundar, se levado (e será) em consideração, o capitalismo e os mercados de vigilância.

É possível caracterizar o *dark data* como dados de uma variedade de tipos, mas essencialmente, dados não estruturados, como mensagens de texto, documentos, e-mails, arquivos de áudio e vídeo (Kambies, Roma, Mittal, & Sharma, 2017). A Gartner (n.d.) caracteriza *dark data* como "ativos de informação que organizações coletam, processam e armazenam durante suas atividades regulares de negócio, mas geralmente falham em utilizá-los para outros fins (como *analytics*, nas relações de negócios ou em direta monetização)" (tradução livre).

Costuma-se comparar o conceito de *dark data* com a *dark matter* ou a *dark energy*. Esta última, por exemplo, compõe cerca de 68% do universo, enquanto a *dark matter* compõe cerca de 27% da matéria existente. Todo o resto da matéria e energia observável, menos de 5%, compõe o restante, sendo aquela que atualmente enxergamos e compreendemos (NASA, 2017). Pode-se associar esse universo observável/não observável à Big Data e tudo aquilo que a ciência de dados consegue/não consegue extrair de informação e conhecimento.

Na realidade, são dois os principais motivos pelos quais o *dark data* é armazenando. O primeiro deles é por uma questão de *compliance*. De grandes a pequenas empresas, frequentemente são armazenados mais dados do que necessário ao estrito funcionamento da sua atividade devido à possibilidade de auditoria (Collet, 2015). O outro motivo será por conta do baixo custo, atualmente, do armazenamento de dados (LaChapelle, 2016), com o avanço dos serviços de *cloud storage*, entre outros, associado à possibilidade de ganhos futuros com a exploração desses dados. Isso justifica o investimento maciço e prioritário em profissionais de *analytics* pelas principais empresas de tecnologia do mundo (Kambies et al., 2017).

O avanço das técnicas de análise de dados vem prometendo uma exploração cada vez mais sofisticada e profunda das cadeias de dados, sugerindo novos níveis de Big Data exploráveis. Recentemente,

pesquisadores da Universidade de Stanford desenvolveram o DeepDive [1], sistema baseado em *machine learning* que promete extrair bases de dados estruturadas a partir de dados não-estruturados (*dark data*) com inigualável acurácia, desencadeando uma disrupção em termos de observação de padrões (Zhang, Shin, Ré, Cafarella, & Niu, 2017) e, consequentemente, comportamentos de usuários e clientes.

Porém esses mesmos mecanismos de análise de dados, baseados em algoritmos cada vez mais complexos e opacos (Pasquale, 2015), inferem informações, associando dados a pessoas, gerando novas formas de rastreamento de comportamento *online*. É um artifício caro aos modelos de negócios baseados em *data-driven* (Skouma & Léonard, 2015), os quais podem comprometer, de forma ainda mais sensível, o direito à privacidade.

No que se refere aos sistemas baseados em inteligência artificial, Shoshana Zuboff (2016) alerta para a equação: primeiro, há um clamor por cada vez maiores e mais amplos usos de serviços, espaços e dispositivos conectados, mantendo seres humanos como fonte primária de dados e "excesso de informações comportamentais"[2]; segundo, aplica-se inteligências artificiais, como *machine learning*, e ciência de dados para o melhoramento contínuo de algoritmos; em terceiro lugar, converte-se as informações comportamentais em produtos de predição, mapeando comportamentos atuais e futuros; por fim, os produtos de predição inauguram novos "mercados para mercados" ou *meta-markets*. São os moldes do funcionamento do capitalismo de vigilância.

Pode-se inferir que os novos níveis de exploração de dados propostos por sistemas como o DeepDive, por exemplo, acendem o alerta para as implicações à privacidade dos usuários diante da exploração de *dark data*. E este cenário também trará consequências se posto face a face com as legislações e regulações atinentes à proteção de dados pessoais.

LEGISLAÇÃO APLICADA E COMPARADA

Optou-se por analisar a legislação de proteção de dados pessoais e disposições gerais existentes nos Estados Unidos, União Europeia e Brasil.

[1] Acesso: http://deepdive.stanford.edu.
[2] No original: "*behavioral surplus*".

Diversos países da União Europeia já previam em seu sistema normativo alguma espécie de proteção à privacidade do indivíduo. Com o advento da Convenção Europeia de Direitos Humanos, em seu Art. 8º, houve um nivelamento porquanto os países viram-se como signatários de um tratado internacional e adaptaram ou incluíram em suas legislações pátrias a garantia ao indivíduo, respeito à sua vida privada e familiar, seu lar e sua correspondência – sujeita a restrições que estejam de acordo com a lei e necessária num estado democrático.

A necessidade de uma uniformização da legislação de proteção de dados pessoais foi alcançada através da Diretiva 95/46/EC, prevendo a obrigatoriedade por parte dos países pertencentes quanto à adoção de medidas legislativas e regulatórias específicas. Muito embora houvesse um marco no que tange a proteção dos dados pessoais, com diversas inovações, em breve ela será substituída pela General Data Protection Regulation - GDPR. A regulação irá afetar frontalmente os modelos de negócio baseados em *data-driven.* As análises mais pessimistas preveem uma redução de 70% nos gastos com publicidade baseada em informações comportamentais (IHS Markit, 2017) e confrontará, consequentemente, as tendências de exploração de *dark data*.

Quanto aos Estados Unidos, do ponto de vista federal, a Quarta Emenda da Constituição garante aos cidadãos o direito de buscas imotivadas. Jurisprudencialmente, esse direito é interpretado de modo que inclua o "direito à privacidade" e o "direito a ser deixado só". No âmbito estadual, a maioria dos estados reconhece o direito ao cidadão de não ter sua vida pessoal invadida.

Não há uma lei federal única que regule a coleta e uso de dados pessoais no âmbito dos Estados Unidos. Há, porém, diversos guias e *frameworks* de “boas práticas” elaborados por agências reguladoras governamentais e de setores da indústria que, muito embora não tenham força de lei, têm ganhado cada vez mais relevância, sendo utilizados de modo coercitivo pelo Estado. Para dados classificados normalmente como sensíveis, a exemplo de informações financeiro-fiscais e de saúde, existem leis específicas. Alguns dispositivos legais especificam quais dados são considerados pessoais e não-públicos, novamente a depender do diploma e sua aplicabilidade. De maneira geral,

quanto à jurisdição, normalmente as leis aplicam-se a companhias e pessoas que façam negócio ou morem nos Estados Unidos.

Vale mencionar as exceções às leis de proteção de dados pessoais, quando por suposto interesse de

segurança pública, ganhando relevo notadamente após os atentados de 11 de setembro de 2001. O *Patriot Act*, substituído posteriormente pelo *Freedom Act*, instituídos sob a justificativa de garantir a segurança nacional, criou enorme celeuma, principalmente após as denúncias de Edward Snowden, que expôs uma série de práticas de vigilância em massa e espionagem governamental.

Recentemente, já no governo de Donald Trump, um enorme retrocesso, reflexo direto da influência do *data-driven market* e das empresas de telecomunicações, várias regulações que tratavam de privacidade e segurança dos usuários de internet foram repelidas, permitindo, por exemplo, provedores de internet comercializar históricos de *browser* e dados de aplicativos do usuário, tendo em vista que a Federal Trade Comission não trata tais dados como sensíveis.

No que tange ao consentimento, os Princípios de Propaganda Comportamental apenas sugerem que obtenham consentimento expresso afirmativo antes de coletarem dados tais como dados financeiros, dados sobre saúde do indivíduo, número do seguro social, entre outros, sendo essa talvez das principais brechas no sistema americano de proteção de dados pessoais. Ressalta-se que o sistema de proteção de dados pessoais norte-americano será afetado pelos impactos decorrentes da entrada em vigor da GDPR europeia.

No âmbito nacional, a Constituição Federal de 1988 prevê em seu Art. 5º a garantia quanto à inviolabilidade da intimidade e vida privada do cidadão, bem como a do sigilo de sua correspondência e comunicações de maneira geral. O que não parece bastante para assegurar a privacidade e proteger os dados pessoais em um cenário e economia de vigilância.

O Marco Civil da Internet, embora inovador e uma referência mundial em termos de legislação referente à internet, com fortes matizes de influência na legislação europeia, fruto de amplo debate público, não avança suficientemente, por apenas resvalar nas questões relativas à proteção de dados pessoais, carecendo inclusive de uma definição assertiva sobre o que seriam dados pessoais. Muito embora já existam de maneira esparsa em nosso ordenamento jurídico previsões sobre o conceito, como na Lei Geral das Telecomunicações, no Código de Defesa do Consumidor, na Lei de Acesso à informação e na Lei da Organização Criminosa, o Marco Civil não é passível de gerar a segurança jurídica necessária (Leite & Lemos, 2014).

Originado no Poder Executivo e submetido à intensa participação popular, há a iminência da Lei

Geral de Proteção de Dados Pessoais Brasileira, constituída no Projeto de Lei nº 5276. Fortemente inspirado nas Diretivas Europeia e Canadense, caso mantenha seu espírito atual, relocará o Brasil em lugar de referência.

A conjugação de uma multidisciplinaridade no que se refere aos diplomas de disposições tangentes à proteção de dados pessoais no Brasil e no mundo é uma necessidade desafiadora, pondo em exercício os aplicadores de normas em um contexto globalizado de economia baseadas, sobretudo, em plataformas online. Os mercados de *data-driven* são diretamente influenciados pelas legislações atuais e pelas que estão por vir, criando um campo de batalha onde a privacidade é o objeto jurídico posto em jogo. Este artigo procura propor que o desenvolvimento de técnicas de exploração de dados, bem como de *dark data*, deve ser observado por este viés, garantindo a inovação de modelos de negócio e de novas descobertas, porém sujeitos à garantia de direitos fundamentais, principalmente aqueles fragilizados em um cenário de mercados de vigilância.

REFERÊNCIAS

Boyd, D., & Crawford, K. (2011, setembro). Six Provocations About Big Data. *A Decade in Internet Time:* Symposium on the Dynamics of the Internet and Society, Oxford, Reino Unido, 01.

Castells, M. (2016). *Sociedade em Rede* [vol. I]. São Paulo: Paz & Terra.

Collet, H. (2015, janeiro 05). Dark Data Management: tohe Next Frontier in Information Governance. *Bloomberg* [versão eletrônica]. Recuperado em 14 de novembro de 2017 de https://www.bloomberg.com/company/announcements/dark-data-management-next-frontier-information-governance.

Desjardins, J. (2017a, maio 12). Chart: Here's How 5 Tech Giants Make Their Billions. *Visual Capitalist*. Recuperado em 15 de novembro de 2017 de http://www.visualcapitalist.com/chart-5-tech-giants-make-billions.

Desjardins, J. (2017b, agosto 13). Chart: The Largest Comanies By Market Cap Over 15 Years. *Visual Capitalist.* Recuperado em 15 de novembro de 2017 de http://www.visualcapitalist.com/chart-largest-companies-market-cap-15-years.

Gartner (n.d.). *IT Glossary:* Dark Data. Recuperado em 14 de novembro de 2017 de https://www.gartner.com/it-glossary/dark-data.

Hernandez, A. (2016, dezembro 11). When it comes to digital advertising, Google and Facebook own your eyes. *Techaeris.* Recuperado em 15 de novembro de 2017 de https://techaeris.com/2016/12/11/digital-advertising-google-facebook.

HIS Markit (2017, setembro). *The Economic Value of Behavioral Targeting in Digital Advertising.* Recuperado em 15 de novembro de 2017 de https://www.iabeurope.eu/wp-content/uploads/2017/09/BehaviouralTargeting_FINAL.pdf.

Kambies, T., Roma, P., Mittal, N., & Sharma, S. K. (2017, fevereiro 07). Dark Analytics: Illuminating Opportunities Hidden Within Unstructured Data. *Deloitte Insights*. Recuperado em 15 de novembro de 2017 de https://dupress.deloitte.com/dup-us-en/focus/tech-trends/2017/dark-data-analyzing-unstructured-data.html.

LaChapelle, C. (2016, março 16). The Cost of Data Storage and Management: Where is It Headed in 2016? *The Data Center Journal*. Recuperado em 13 de novembro de 2017 de http://www.datacenterjournal.com/cost-data-storage-management-headed-2016.

Leite, G., & Lemos, R. (Org.) (2014). *Marco Civil da Internet*. São Paulo: Atlas.

Skouma, G., & Léonard, L. (2015). On-line Behavioral Tracking: What May Change After the Legal Reform on Personal Data Protection. In. S. Gutwirth, R. Leenes, & P. Hert (Eds.). *Law, Governance and Technology Series* (Vol. 20, Cap. 2, pp. 35-60). Nova Iorque: Springer.

NASA, National Aeronautics and Space Administration (2017). *Dark Energy, Dark Matter*. Recuperado em 13 de novembro de 2017 de https://science.nasa.gov/astrophysics/focus-areas/what-is-dark-energy.

Pal, K. (2017). What is the importance of dark data in Big Data world? *KDNuggets News*, *15*(39) [versão eletrônica]. Recuperado em 14 de novembro de 2017 de https://www.kdnuggets.com/2015/11/importance-dark-data-big-data-world.html.

Pasquale, F. (Org.). (2015). *The Black Box Society*. Cambridge: Harvard University Press.

The Economist (2017, maio 06). *The World Most Valuable Resource is no Longer Oil, but Data.* Recuperado em 15 de novembro de 2017 de https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource.

Zhang, C., Shin, J., Ré, C., Cafarella, M., & Niu, F. (2017, maio). Extracting Databases from Dark Data with DeepDive. *The ACM SIGMOD International Conference on Management of Data* (SIGMOD), Chicago, EUA, 43.

Zuboff, S. (2015). Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology, 30*, 75-89. Recuperado em 15 de novembro de 2017 de http://papers.ssrn.com/abstract=2594754.

Zuboff, S. (2016). The Secrets of Surveillance Capitalism. *Frankfurter Allgemeine Zeitung* [versão eletrônica]. Recuperado em 14 de novembro de 2017 de http://www.faz.net/aktuell/feuilleton/debatten/the-digital-debate/shoshana-zuboff-secrets-of-surveillance-capitalism-14103616.html?printPagedArticle=true#pageIndex_2.