

**CONGRESSO INTERNACIONAL DE
DIREITO E INTELIGÊNCIA
ARTIFICIAL**

**INTELIGÊNCIA ARTIFICIAL E TECNOLOGIAS
APLICADAS AO DIREITO IV**

ANA CAROLINA REIS PAES LEME

I61

Inteligência artificial e tecnologias aplicadas ao direito IV [Recurso eletrônico on-line]
organização Congresso Internacional de Direito e Inteligência Artificial: Skema Business
School – Belo Horizonte;

Coordenadores: Leonardo Vieira Wandelli, Ana Carolina Reis Paes Leme e José Eduardo
Chaves Júnior – Belo Horizonte: Skema Business School, 2020.

Inclui bibliografia

ISBN: 978-65-5648-103-6

Modo de acesso: www.conpedi.org.br em publicações

Tema: Desafios da adoção da inteligência artificial no campo jurídico.

1. Direito. 2. Inteligência Artificial. 3. Tecnologia. I. Congresso Internacional de Direito
e Inteligência Artificial (1:2020 : Belo Horizonte, MG).

CDU: 34



CONGRESSO INTERNACIONAL DE DIREITO E INTELIGÊNCIA ARTIFICIAL

INTELIGÊNCIA ARTIFICIAL E TECNOLOGIAS APLICADAS AO DIREITO IV

Apresentação

É com enorme alegria que a SKEMA Business School e o CONPEDI – Conselho Nacional de Pesquisa e Pós-graduação em Direito apresentam à comunidade científica os 14 livros produzidos a partir dos Grupos de Trabalho do I Congresso Internacional de Direito e Inteligência Artificial. As discussões ocorreram em ambiente virtual ao longo dos dias 02 e 03 de julho de 2020, dentro da programação que contou com grandes nomes nacionais e internacionais da área, além de 480 pesquisadoras e pesquisadores inscritos no total. Estes livros compõem o produto final deste que já nasce como o maior evento científico de Direito e da Tecnologia do Brasil.

Trata-se de coletânea composta pelos 236 trabalhos aprovados e que atingiram nota mínima de aprovação, sendo que também foram submetidos ao processo denominado double blind peer review (dupla avaliação cega por pares) dentro da plataforma PublicaDireito, que é mantida pelo CONPEDI. Os quatro Grupos de Trabalho originais, diante da grande demanda, se transformaram em 14 e contaram com a participação de pesquisadores de 17 Estados da federação brasileira. São cerca de 1.500 páginas de produção científica relacionadas ao que há de mais novo e relevante em termos de discussão acadêmica sobre os temas Direitos Humanos na era tecnológica, inteligência artificial e tecnologias aplicadas ao Direito, governança sustentável e formas tecnológicas de solução de conflitos.

Os referidos Grupos de Trabalho contaram, ainda, com a contribuição de 41 proeminentes professoras e professores ligados a renomadas instituições de ensino superior do país, os quais indicaram os caminhos para o aperfeiçoamento dos trabalhos dos autores. Cada livro desta coletânea foi organizado, preparado e assinado pelos professores que coordenaram cada grupo. Sem dúvida, houve uma troca intensa de saberes e a produção de conhecimento de alto nível foi, certamente, o grande legado do evento.

Neste norte, a coletânea que ora torna-se pública é de inegável valor científico. Pretende-se, com esta publicação, contribuir com a ciência jurídica e fomentar o aprofundamento da relação entre a graduação e a pós-graduação, seguindo as diretrizes oficiais. Fomentou-se, ainda, a formação de novos pesquisadores na seara interdisciplinar entre o Direito e os vários campos da tecnologia, notadamente o da ciência da informação, haja vista o expressivo

número de graduandos que participaram efetivamente, com o devido protagonismo, das atividades.

A SKEMA Business School é entidade francesa sem fins lucrativos, com estrutura multicampi em cinco países de continentes diferentes (França, EUA, China, Brasil e África do Sul) e com três importantes creditações internacionais (AMBA, EQUIS e AACSB), que demonstram sua vocação para ensino e pesquisa de excelência no universo da economia do conhecimento. A SKEMA, cujo nome é um acrônimo significa School of Knowledge Economy and Management, acredita, mais do que nunca, que um mundo digital necessita de uma abordagem transdisciplinar.

Agradecemos a participação de todos neste grandioso evento e convidamos a comunidade científica a conhecer nossos projetos no campo do Direito e da tecnologia. Já está em funcionamento o projeto Nanodegrees, um conjunto de cursos práticos e avançados, de curta duração, acessíveis aos estudantes tanto de graduação, quanto de pós-graduação. Até 2021, será lançada a pioneira pós-graduação lato sensu de Direito e Inteligência Artificial, com destacados professores da área.

Agradecemos ainda a todas as pesquisadoras e pesquisadores pela inestimável contribuição e desejamos a todos uma ótima e proveitosa leitura!

Belo Horizonte-MG, 07 de agosto de 2020.

Profª. Drª. Geneviève Daniele Lucienne Dutrait Poulingue

Reitora – SKEMA Business School - Campus Belo Horizonte

Prof. Dr. Edgar Gastón Jacobs

Coordenador Acadêmico da Pós-graduação de Direito e Inteligência Artificial da SKEMA Business School

PRÁTICA DE WEBSCRAPING EM BASES DE DADOS PÚBLICAS

PRACTICE OF WEBSCRAPING IN PUBLIC DATABASES

Mariana Ribeiro França Guerra Magalhães

Robert Emmanuel de Oliveira

Resumo

Este trabalho tem por objetivo analisar tecnologias como o Big Data e a utilização de Web Scraping para tratamento de dados em bases de dados públicas, utilizando uma pesquisa empírica e com métodos qualitativos indutivos. Serão abordados a evolução da sociedade até a Quarta Revolução Industrial, a economia do conhecimento e a sociedade da informação. Ademais, o presente trabalho abordará as tendências de dados abertos mundiais que influenciaram o Brasil em matéria de disponibilização de dados de bases públicas visando a transparência da gestão pública e a prática da raspagem de dados pessoais nessas bases de dados.

Palavras-chave: Sociedade da informação, Dados abertos, Privacidade, Proteção de dados, Web scraping

Abstract/Resumen/Résumé

This work aims to analyze technologies such as Big Data and the use of Web Scraping to collect and process data in public databases, using empirical research and with inductive qualitative methods. The evolution of society until the Fourth Industrial Revolution will be addressed, when we will enter into the concepts of knowledge economy and information society. In addition, this paper will address trends in open world data that have also influenced Brazil in terms of making public data available with a view to transparency in public management and the practice of scraping personal data in these databases.

Keywords/Palabras-claves/Mots-clés: Information society, Open data, Privacy, Data protection, Web scraping

1. INTRODUÇÃO

Até chegarmos a uma sociedade pautada em dados e informações, diversas foram as mudanças nas formas de organização e estruturação social e econômica. Há mais de 10.000 anos, viveu-se a primeira revolução, conhecida por revolução agrícola, em que a atividade e o produto agrícola eram os elementos centrais que movimentavam a economia; em seguida, ocorreram a Primeira e Segunda Revoluções Industriais, em que, a partir da invenção de máquinas e da descoberta e utilização de fontes de energia, como a elétrica, o ponto central da economia migrou-se para a indústria e para a produção fabril; posteriormente, passamos por outra Revolução Industrial, que incluiu a Computação, Automação e outras fontes de energia à indústria; e, por fim, antes mesmo de ter início a Quarta Revolução Industrial, na sociedade pós-industrial, a prestação de serviços assumiu o protagonismo enquanto atividade de impulso à economia (SCHWAB, 2016).

Na era pós-industrial, a informação se tornou o principal elemento de desenvolvimento da economia, de modo que recursos que outrora foram protagonistas nas sociedades agrícola e industrial migraram para um segundo plano (BIONI, 2018). Como consequência, a informação se tornou o capital, de modo que deixamos de ter o capital direcionado para ativos tangíveis e migramos para ativos intangíveis, o que tornou possível exponenciar retornos em escala, criando riquezas com muito menos força de trabalho, estrutura industrial e custos do que anteriormente.

Para efeito de comparação, consideremos a cidade de Detroit, na década de 90, polo industrial dos EUA. As três maiores empresas instaladas na região possuíam uma capitalização de US\$36 bilhões, faturamento de US\$250 milhões, e 1.2 milhão de empregados. Em 2014, as três maiores empresas instaladas no Vale do Silício tinham capitalização de mercado de cerca de US\$1 trilhão, obtinham receitas na mesma casa daquelas em Detroit, mas contavam com 10 vezes menos empregados que a região industrial da década de 90. (SCHWAB, 2016)

Há muito tempo informação e poder são conhecidos por andar juntos. Trazendo essa perspectiva para a realidade atual, podemos dizer que a ampla disseminação da informação reflete um poder desconcentrado, que resulta em elevados níveis de transparência. São óbvias, portanto, as razões pelas quais a *transparência* tem se mostrado tão presente na atualidade. O poder (informação) não está mais nas mãos de alguns; ele está disponível para acesso de todos. A informação é atualmente o elemento que estrutura toda a nossa sociedade, logo, não pode ser

surpresa que a transparência tenha tomado uma posição de destaque neste cenário, inclusive em relação às bases de dados públicas, que serão tratadas em sequência.

O objetivo do presente texto é demonstrar como as políticas de dados abertos têm sido implementadas no Brasil; analisar o que dispõe o ordenamento jurídico brasileiro no que diz respeito ao tratamento de dados disponibilizados em bases de dados públicas; e, por fim, examinar os impactos e a possibilidade de conciliação das novas tecnologias com as atividades de tratamento de dados, ou, mais do que isso, a possibilidade de haver raspagem de dados em bases de dados públicas.

2. ACESSO À INFORMAÇÃO E BASES DE DADOS PÚBLICAS

Na gestão das organizações públicas, temos uma geração imensa de dados de todos os tipos, inclusive dados pessoais. E, para alcançar a transparência almejada pelas políticas de dados abertos, não há outro caminho que não seja a disponibilização desses dados de forma acessível a tecnologias de análise de dados.

Países como EUA e Inglaterra já caminham nesse sentido desde 2009, assim como o Brasil, que também tem buscado aplicar a transparência nos dados públicos. Em 2011, após um período de discussões, constituiu-se a LAI (Lei de acesso a informação, nº12.527/2011). A partir daí, tivemos iniciativas de todos os órgãos e empresas públicas em prol da disponibilização de dados. Políticas de dados abertos e disponibilização de dados de organizações públicas são tendências mundiais que visam aumentar a participação do cidadão e trazer transparência para a gestão pública, iniciando uma batalha contra a corrupção e fraudes. (SANTAREM SEGUNDO, 2013)

Antes disso, um movimento de busca por dados abertos segue propósitos e objetivos voltados para a economia digital e descentralização do poder que os dados representam na sociedade da era pós-industrial. Organizações como a Open Knowledge Foundation fomentam um movimento que tem como pauta central a dificuldade estrutural enfrentada para conter desigualdades na economia digital. (ABRAMOVAY; ZANATTA, 2019)

É certo que dados abertos significam transparência; mas isso não significa que a agenda de busca por um mundo de dados abertos e disponíveis não reconheça, também, a importância de se proteger dados pessoais, apesar de a principal preocupação deste movimento ser a não permissão às empresas baseadas em tecnologias pautadas em dados consolidarem posições extremamente

monopolistas em setores estratégicos. Por mais que, na sociedade moderna, o conceito de privacidade tenha evoluído, não podemos ignorar os riscos e preocupações que surgiram em decorrência do crescimento exponencial de empresas que têm dados pessoais como principal ativo e moeda de troca por produtos e serviços, e da utilização de tecnologias como *Big Data* e *Data Scraping*.

O Brasil foi pioneiro em tratar sobre os dados abertos, mas nunca teve uma cultura voltada para preocupação com a proteção de dados, e isso fica claro com o nascimento da LAI, que foi uma das primeiras legislações do mundo a incorporar políticas de dados abertos. Seguindo as tendências internacionais, a LAI adota os princípios de divulgação máxima e limitação das exceções, ou seja, o acesso total é a regra, enquanto o sigilo é a exceção. (POSSAMAI; GONZATTI DE SOUZA, 2020)

Vale destacar que, em matéria de dados pessoais cujo acesso é público, é perceptível, por parte daqueles que realizam o tratamento dos dados, pouca ou nenhuma preocupação quanto à coleta e utilização dos mesmos. Essa falta de cuidado é ocasionada pela falsa ideia que muitos possuem de que, se o dado é público, não há qualquer restrição quanto ao seu uso, podendo ele ser usado livremente.

Contudo, esta é uma visão completamente equivocada das políticas de acesso a dados e das bases de dados públicas, já que, além de outras legislações como o Código de Defesa do Consumidor (Lei n. 8.078/1990) e o Marco Civil da Internet (Lei n. 12.965/2014), a LAI também aborda restrições ao acesso a dados de natureza pessoal, ou seja, dados que identificam pessoas naturais. Em linhas gerais, a limitação está diretamente atrelada àqueles dados que se referem à intimidade, à vida privada, à honra e à imagem do titular (Art.31), não sendo permitido, nestes casos, o acesso de terceiros sem o consentimento do titular.

O consentimento do titular dos dados foi a máxima utilizada em todas as legislações que tratam sobre dados pessoais no Brasil. E, mesmo com o nascimento de legislações por todo o mundo, específicas para o tratamento de dados pessoais, o titular dos dados pessoais continua sendo o ponto focal. (BIONI, 2018)

Com a vigência da Lei Geral de Proteção de Dados (LGPD), publicada em 2018, o acesso ou tratamento de dados pessoais disponibilizados em bases de dados públicas passa a ter uma definição mais objetiva, sendo possível traçar um caminho mais claro sobre quais são os tipos de

dados restritos e como eles podem ser utilizados, inclusive sem a necessidade do consentimento do titular.

Com a evolução e o desenvolvimento de tecnologias como o *Big Data* e o *Data Scraping*, surgem questionamento como: quais são os limites da utilização de tecnologias como o *Big Data* neste cenário? Será legalmente possível a raspagem de dados em bases de dados públicas?

3. WEB SCRAPING EM BASES DE DADOS PÚBLICAS

De uma simples leitura e interpretação dos conceitos de *Big Data* e *Data Scraping*, ou raspagem de dados, é possível perceber que a evolução das novas tecnologias tem percorrido um caminho que exige a realização de modificações legislativas para que possua aplicabilidade prática. Do contrário, torna-se inviável, já que a abordagem normativa que imperou até aqui é extremamente rígida, restritiva e pouco ou nada flexível.

Big Data não é exatamente uma tecnologia, ou um sistema, o termo se refere, ao mesmo tempo, à quantidade de dados disponíveis no ambiente digital e à possibilidade, por meio de métodos, técnicas e diversos sistemas e algoritmos disponíveis, do tratamento e transformação de dados estruturados e desestruturados em conhecimento. “O termo *Big Data* surgiu para definir arquiteturas de sistemas capazes de lidar com as novas dimensões dos dados: velocidade, variedade e volume” (AZEVEDO; NEVES; NOVO, 2014).

Data Scraping ou *Web Scraping* também não é uma tecnologia em si, mas a atividade de se extrair, de forma automatizada, dados em páginas da internet. É como um programa capaz de consultar um servidor Web, solicitar dados (como se fosse uma página WEB) e extrair informações ao analisá-lo. O termo, portanto, refere-se à técnica de coletar dados de páginas web por meio de uma Application Program Interface (API). (MITCHELL, 2016)

A produção em massa de dados no ambiente digital e a popularização do *Data Scraping* têm feito com que o controle relacionado a dados acessíveis em bases de dados públicas seja muito dificultado. Por outro lado, a utilização desse tipo de atividade, por si só, não significa a ocorrência de violação à privacidade ou a dados de pessoas naturais.

Bases de dados públicas são criadas e alimentadas com finalidades definidas e também são disponibilizadas, conforme política de dados abertos, com finalidades definidas. Além disso, não há como nos relacionar em sociedade sem entregar dados, seja para empresas privadas, seja para

organizações e empresas públicas. Em praticamente todas as atividades cotidianas que realizamos em sociedade, a disponibilização de dados mostra-se elemento crucial para a identificação, segurança, fornecimento de benefícios e serviços, e compra de produtos, e tudo isso, como já pontuado, não só entre indivíduo e Estado, mas também em relação a empresas privadas e demais indivíduos.

Apesar da necessidade de garantia da privacidade e da segurança de dados pessoais, é praticamente impossível que todos os dados que identificam uma pessoa estejam completamente fora do acesso de terceiros. As proteções que circundam dados pessoais e privacidade são fundamentais não para que os dados não sejam acessados, mas para que a utilização de dados pessoais aconteça de forma comedida; que o titular possua o máximo controle possível de seus dados; que seus direitos fundamentais sejam preservados; e que o acesso a dados não possa representar discriminação desse indivíduo.

Nesse mesmo sentido, Danilo Doneda (2006) diz que:

A utilização de dados pessoais em diversas atividades não é, em si, um problema. Na verdade ela torna possíveis certas empreitadas com um alto grau de eficiência, em áreas que vão do planejamento administrativo à pesquisa de mercado. Ocorre que esta atividade requer instrumentos que possibilitem aos interessados um efetivo controle em relação aos seus dados pessoais, garantindo acesso, a veracidade, a segurança, o conhecimento da finalidade para o qual serão utilizados (entre outros). (DONEDA, 2006, p. 2)

Nesse sentido, a problemática envolvendo a coleta e tratamento de dados por *Data Scraping* não circunda a própria atividade, mas os motivos e justificativas para que determinadas tecnologias serão utilizadas possibilitando a coleta de dados pessoais.

Até aqui, por inúmeras vezes foi abordado o fato de que o consentimento sempre foi tido como o elemento central para o tratamento de dados pessoais. No entanto, esta tradição normativa tornou-se incompatível com a sociedade contemporânea, que apresenta-se dinâmica em todos os seus aspectos, inclusive no que se refere ao elevado fluxo de dados pessoais. A produção de dados no mundo digital sofreu um aumento exponencial, como já evidenciado inúmeras vezes até aqui, e exigir dos controladores e operadores de dados que obtenham o consentimento dos titulares a todo momento poderá os levar a um estado de exaustão, inviabilizando as atividades de tratamento de dados e o desenvolvimento de novas tecnologias (BIONI, 2020).

Conforme dispõe o artigo 7º, §3º, para que seja lícito o tratamento de dados públicos ou tornados públicos pelo titular, deve-se considerar a finalidade, a boa fé e o interesse público que

justificaram a disponibilização daquele dado, sendo ainda possível o tratamento para novas finalidades, nos termos do §7º, desde que "observados os propósitos legítimos e específicos para o novo tratamento e a preservação dos direitos do titular" (LGPD, 2018).

Não há, na LGPD, uma definição expressa do que seriam propósitos legítimos. Entretanto, pela experiência que têm sido vivenciada em outros países, especialmente da Europa, onde vigora o Regulamento Geral de Proteção de Dados (GDPR), este propósito legítimo é identificado a partir de uma análise contextual em cada caso concreto.

Existem interpretações contemporâneas de que, por trás dos propósitos legítimos, haverá uma flexibilidade maior ao tratamento de dados. Estas interpretações residem nas razões da emenda parlamentar que levou à criação do referido §7º, por meio das quais o relator reconheceu que, quando publicamente acessíveis, os dados pessoais passam a ser, também, um importante elemento para a realização de análises e estudos, "promovendo competitividade, inovação, empregabilidade e prosperidade". (BASTOS; et.al, 2019)

Além das definições trazidas com a LGPD no que tange ao tratamento de dados de bases públicas para finalidades diversas daquelas às quais os dados foram disponibilizados, é importante ressaltar que a legislação traz, ainda, outras nove bases legais, além do consentimento do titular, para o tratamento de dados pessoais. O cumprimento de uma relação contratual, por exemplo, poderia tornar legal o acesso a dados pessoais; o exercício regular de direitos; ou a proteção ao crédito.

Sob essa perspectiva, portanto, o tratamento de dados de bases públicas, por exemplo, por pessoas jurídicas, que tenha por finalidade a proteção ao crédito deve ser admitido, mesmo que por meio da utilização de *Web Scraping* já que encontra-se em consonância com uma das bases legais da nova Lei.

4. CONCLUSÃO

A evolução e o desenvolvimento das novas tecnologias que possibilitam o tratamento de dados de forma automatizada, como o *Big Data* e o *Data Scraping*, têm ocupado cada vez mais espaço na sociedade contemporânea. Neste contexto, surge uma preocupação atinente à privacidade dos titulares dos dados, direito fundamental garantido pelo ordenamento jurídico brasileiro, e aos reflexos desta nas atividades de tratamento de dados.

É certo que o tratamento de dados deve sempre ser feito com muita cautela, mas o direito à privacidade dos titulares, por si só, não o proíbe. Com a vigência da LGPD, o esperado é que seja aceita a possibilidade de tratamento de dados, inclusive por meio da utilização deste tipo de técnicas, desde que os controladores e operadores de dados estejam sob o amparo de uma das bases legais que a nova Lei nos traz, e que já foram mencionadas nesta pesquisa; ou desde que realizem o tratamento dos dados para finalidade compatível àquela que ensejou a disponibilização dos mesmos.

BIBLIOGRAFIA

ABRAMOVAY, Ricardo; ZANATTA, Rafael Augusto Ferreira. *Dados Pessoais Abertos: Pilares dos Novos Mercados Digitais?*. RDU: Porto Alegre, Volume 16, n. 90, 2019, 155-178, nov-dez 2019.

AZEVEDO, M.M.; NEVES, J.M.S.; NOVO, R.F. *O crescimento do Big Data e as possíveis implicações éticas do seu uso na análise das redes sociais*. In: WORKSHOP DE PÓS-GRADUAÇÃO E PESQUISA DO CENTRO PAULA SOUZA, 9., Estratégias Globais e Sistemas Produtivos Brasileiros, 2014

BASTOS, Rodrigo Albero Caldeira; SCHVARTZMAN, Felipe; PIERI, José Eduardo de V. *Dados pessoais 'públicos' são, de fato, públicos?*. [2019] Disponível em: <<https://www.jota.info/opiniao-e-analise/artigos/dados-pessoais-publicos-sao-de-fato-publicos-30062019>> Acesso em: 18 de junho de 2020.

BIONI, BRUNO RICARDO. *Proteção de Dados Pessoais - A Função e os Limites do Consentimento*. 1. ed. Rio de Janeiro: Forense, 2018.

DONEDA, Danilo. *Da privacidade à proteção de dados pessoais*. Rio de Janeiro: Renovar, 2006.

MITCHELL, R. *Web Scraping com Python*. São Paulo: Novatec, 2016.

POSSAMAI, Ana Júlia; GONZATTI DE SOUZA, Vitoria. *Transparência e Dados Abertos Governamentais: Possibilidades e Desafios a Partir da Lei De Acesso à Informação*. Revista Administração Pública e Gestão Social, vol. 12, núm. 2, 2020. Disponível em: <<http://www.redalyc.org/articulo.oa?id=351562414012>> Acesso em: 18 de junho de 2020.

SANTAREM SEGUNDO, J. E. *Tecnologías de la información y la comunicación para proporcionar datos abiertos en formato semántico*. Ibersid, v. 7, p.33-40, 2013. Disponível em: <<http://ibersid.eu/ojs/index.php/ibersid/article/view/4075/3744>> Acesso em: 17 de junho de 2020.

SCHWAB, Klaus. *A Quarta Revolução Industrial*. São Paulo: Edipro, 2018.