

**BIG DATA ANALYTICS E O CIENTISTA SOCIAL.****BIG DATA ANALYTICS AND THE SOCIAL SCIENTIST.**

Oswaldo Soulé Junior\*  
Vanderlei de Castro Ezequiel\*\*  
Maria Teresa Miceli Kerbauy\*\*\*

**RESUMO**

O mundo observa um crescimento exponencial no volume de dados gerados nas mais variadas fontes e a altíssima velocidade. A isso denominou de *Big Data*. Os paradigmas da TIC de Big Data são o Hadoop e NoSQL, o que exige novas expertises para profissionais que atuam nessa área. O profissional de *Big Data* é o *Data Scientist*, precisa de ampla variedade de conhecimentos para a sua atuação. Este trabalho busca trazer a compreensão do que é o *Big Data*, o profissional que atua nessa área, o *Data Scientist* e os desafios do profissional de Ciências Sociais frente a esse ambiente. Através dessa pesquisa, foi possível verificar que esse é um novo e amplo campo que se abre ao Cientista Social, mas ao mesmo tempo, pode fazer com que ele sofra a concorrência de profissões vizinhas por um mercado de trabalho cada vez mais concorrido, mais desregulamentado, precário e tecnológico.

**Palavras-chave:** Big Data; Ciências Sociais; Analytics; Cientista Social;

**ABSTRACT**

The world is seeing an exponential growth in the volume of data generated from the most varied sources and at very high speeds. This is called Big Data. Big Data paradigms are Hadoop and NoSQL, which requires new expertise for professionals working in this area. The Big Data professional is the Data Scientist, he needs a wide range of knowledge for his performance. This paper seeks to bring understanding of what is Big Data, who is the professional working in this area, the Data Scientist and the challenges of the Social Sciences professional facing this environment. Through this research, it was possible to verify that this is a new and wide field that is open to the Social Scientist, but at the same time, it can make it suffer competition from neighboring professions for an increasingly competitive, more deregulated labor market. precarious and technological.

**Keywords:** Big Data; Social Science; Analytics; Data Scientist.

---

\*Doutorando - Universidade Estadual Paulista "Júlio de Mesquita Filho" – Unesp, Araraquara/SP

\*\*Doutorando - Universidade Estadual Paulista "Júlio de Mesquita Filho" – Unesp, Araraquara/SP

\*\*\*Doutora - Universidade Estadual Paulista "Júlio de Mesquita Filho" – Unesp, Araraquara/SP

## INTRODUÇÃO.

A cada ano que passa o mundo observa o crescimento vertiginoso do volume de dados. Segundo Giddens(1991, p. 8) :

...muita gente argumenta que estamos no limiar de uma nova era, a qual as ciências sociais devem responder e que está nos levando para além da própria modernidade. Uma estonteante variedade de termos tem sido sugerida para esta transição, alguns dos quais se referem positivamente à emergência de um novo tipo de sistema social (tal como a “sociedade de informação” ou a “sociedade de consumo”), mas cuja maioria sugere que, mais que um estado de coisas precedente, está chegando a um encerramento (“pós-modernidade”, “pós-modernismo”, “sociedade pós-industrial”, e assim por diante). Alguns dos debates sobre estas questões se concentram principalmente sobre transformações institucionais, particularmente as que sugerem que estamos nos deslocando de um sistema baseado na manufatura de bens materiais para outro relacionado mais centralmente com informação.

O crescimento do volume de dados está totalmente associado a evolução dos recursos da Tecnologia da Informação e Comunicação, as chamadas TIC. Dentre os diversos recursos da TIC, o que gerou o maior volume de dados foi o advento da Internet. Associado a isso, o aumento de dispositivos ou “coisas” conectados à Internet, o qual denominou-se *Internet of Things* – IoT, veio acelerar esse crescimento. Segundo Evans (2011, p. 3), o Cisco Internet Business Solutions Group - IBSG, em 2011, determinou como sendo surgimento do IoT o momento em que o número de dispositivos ou “coisas” conectadas a Internet ultrapassou o número de pessoas conectadas. Fato que se deu entre 2008 e 2009.

Esse grande volume de dados deu origem ao que se acostuma chamar de *Big Data*, termo cunhado por Doug Laney, em 2001, (SICULAR, 2013, p. 1), para identificar a captura, armazenamento, análise e interpretação de um grande volume de dados. A análise desses dados, com o objetivo de obter valor para os negócios, recebeu o nome de *Big Data Analytics*. Este tipo de análise permite que se trabalhe com a análise de um grande volume de dados, obtidos em tempo real ou não. Isto permite que as análises tragam resultados que possam prever determinados comportamentos, percepções e gerar *insights*.

Trazendo o foco para as Ciências Sociais, essas análises podem permitir ao cientista social uma maior compreensão do que ocorre na sociedade, na política e nos mais variados ambientes de encontros das pessoas. Isso demanda novas expertises que devem ser adquiridas por esse profissional.

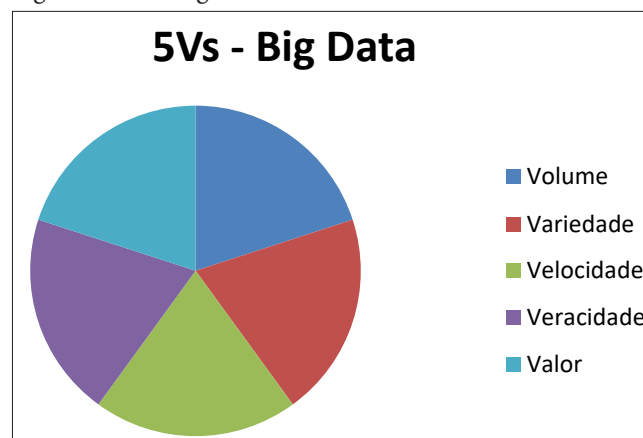
Este trabalho busca trazer a compreensão do que é o *Big Data Analytics* e o profissional que atua nessa área, o *Data Scientist* e, a partir dessa compreensão, mostrar que o profissional de Ciências Sociais e das Ciências Sociais Aplicadas tem que integrar o *Big Data Analytics* em suas atividades para aprimorar as suas atividades e manter-se competitivo em um mercado de trabalho cada vez mais desregularizado, precário e disputado.

A metodologia utilizada para a elaboração deste trabalho parte de uma revisão bibliográfica sobre o tema *Big Data Analytics*, sobre o profissional *Data Scientist* e o sobre profissional de Ciências Sociais; em um primeiro momento traz os conceitos que envolvem o *Big Data Analytics*, em um segundo momento traz as características dos profissionais que atuam nessa área, os *Data Scientists*. Por fim, faz uma reflexão sobre os impactos desse ambiente junto aos profissionais das Ciências Sociais e das Ciências Sociais Aplicadas.

## BIG DATA ANALYTICS.

A definição de *Big Data* envolve o que se costuma chamar de 5 Vs, Figura 1.

Figura 1 – 5Vs Big Data



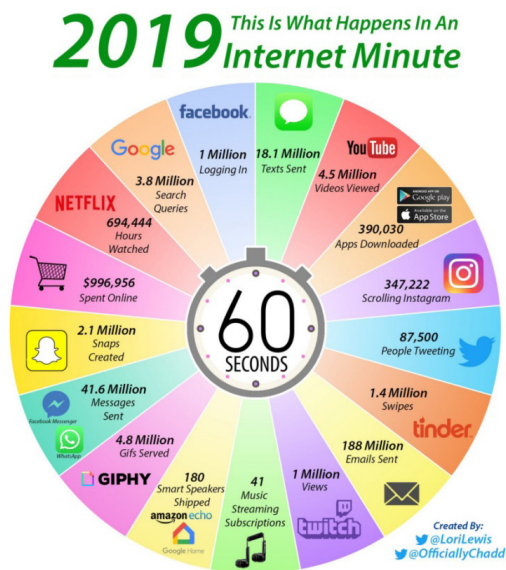
Elaboração própria.

Para que ocorra o *Big Data* é necessário um grande Volume de dados, que podem ter sua origem em uma grande Variedade de fontes, que possam ser

processados e analisados em grande Velocidade, que possam ter a sua Veracidade comprovada e que, por fim, gerem Valor para o negócio. (ISHWARAPPA, ANURADHA, 2015, p. 320)

O grande volume de dados que compõe o Big Data, em alta velocidade, tem sua origem em diversas fontes. Na Figura 2, pode-se observar a origem e o volume de dados gerado em apenas 60 segundos na Internet, o qual demonstra a profusão de fontes que oferecem uma gama de informações, dos mais variados tipos, que atingem as pessoas em seu dia a dia.

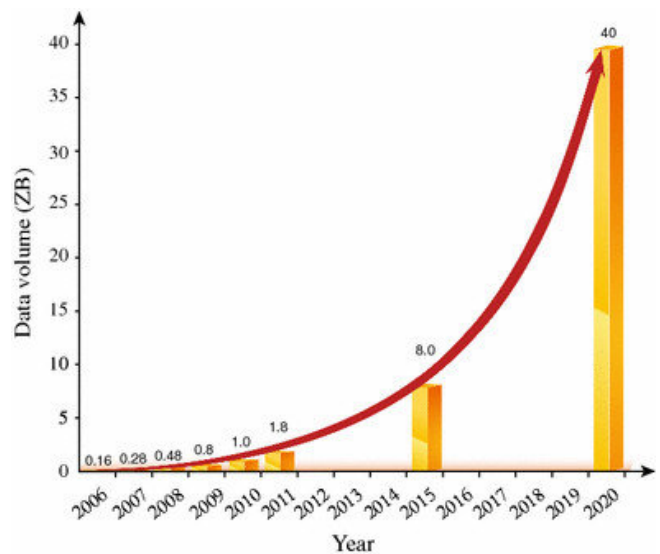
Figura 2 – What happens in a Internet Minute in 2019?



Fonte: <https://www.visualcapitalist.com/what-happens-in-an-internet-minute-in-2019/>

Esse ambiente complexo e variado provoca, como mostra a figura 3, um crescimento exponencial no volume da dados disponíveis, saindo de um valor 0,16 ZB em 2006, para um valor estimado de 40ZB<sup>1</sup> de dados em 2020.

Figura 3 – Evolução do Volume de Dados entre os anos de 2006 e 2020.



Fonte: ReseachGate.

Ao trabalhar com análise desse grande volume de dados, o *Big Data Analytics*, as organizações buscam informações que permitam prever mudanças e qual a reação mais adequada perante a essas mudanças, bem como, identificar possíveis melhorias e *insights* para os seus negócios. O *Big Data Analytics* é a melhor maneira de descobrir novos segmentos de clientes, identificar os melhores fornecedores, associar afinidades entre produtos, entender as vendas, a sazonalidade e assim por diante. (RUSSOM, 201, p. 5, tradução própria).

Para McAfee, Brynjolfsson (2012, p. 4) empresas de venda online podem:

rastrear não apenas o que os clientes compraram, mas também o que mais eles olharam, como eles navegaram pelo site; como eles foram ou não influenciados por promoções, críticas e layouts de página; perceber semelhanças entre os indivíduos e grupos. Em pouco tempo, elas desenvolveram algoritmos para prever quais livros clientes individuais gostariam de ler a seguir - algoritmos que tiveram melhor desempenho cada vez que o cliente respondeu ou ignorou uma recomendação. Varejistas tradicionais simplesmente não conseguem acessar esse tipo de informação, muito menos agir sobre isso em tempo hábil. Não é de admirar que a Amazon colocou tantas livrarias de tijolo e argamassa fora de o negócio. (Tradução própria).

Para se obter os melhores resultados do *Big Data Analytics* é necessário mais e melhores dados.

<sup>1</sup> ZetaByte(ZB) é uma unidade de medida em TI e representa 1.180.591.620.717.411.303.424 (2<sup>70</sup>) Bytes ou 1.099.511.627.776 GB.

Para compor essa base de dados, pode-se integrar diferentes tipos de dados, sendo eles dados estruturados, semiestruturados e não estruturados. A partir das definições de Praveen e Chandra (2017, p.68-69), pode-se estabelecer que:

- Dados estruturados; são aqueles advindos das formas tradicionais de armazenamento e de manipulação que possuem uma estrutura pré-definida e regular de representação previamente definida, tais como Sistemas Gerenciadores de Banco de Dados - SGBDs, arquivos de planilha eletrônica, etc;
- Dados semiestruturados; são aqueles que não são totalmente estruturados, mas também não são desestruturados. Possuindo uma estrutura irregular, embutida no próprio dado e nem sempre possuiu um esquema definido, como páginas da Web, documentos de texto, e-mail, etc;
- Dados não estruturados; são aqueles que não se consegue identificar uma organização clara da sua estrutura, como os arquivos de vídeo e áudio, etc.

O grande repositório onde esse imenso volume de dados, em sua forma bruta, é armazenado foi denominado por James Dixon, *Chief Technical Officer* - CTO da Pentaho como *Data Lake*. Para Marr(2018, p.1),

Um Data Lake contém dados de forma não estruturada e não há uma hierarquia ou organização entre os dados individuais. Ele contém dados em sua forma mais bruta - não é processado ou analisado. Além disso, os Data Lakes aceitam e retêm todos os dados de todas as fontes de dados, suportam todos os tipos de dados e esquemas (a maneira como os dados são armazenados em um banco de dados) são aplicados somente quando os dados estão prontos para serem usados. (Tradução própria).

Os dois grandes paradigmas do *Big Data Analytics* são Hadoop e o NoSQL. Hadoop é uma estrutura *open source*<sup>2</sup>, desenvolvida pela Apache, usada para armazenar e processar grande volume de dados. A AWS-Amazon (SD), explica que o Hadoop possui quatro módulos centrais:

- *Hadoop Distributed File System* (HDFS) - Sistema de arquivos distri-

buído que é executado em hardware padrão ou *low-end*<sup>3</sup>. O HDFS fornece melhor rendimento de dados que os sistemas de arquivos tradicionais, além de alta tolerância a falhas e suporte nativo a grandes conjuntos de dados;

- *Resource Negotiator* (YARN) - Gerencia e monitora nós de cluster<sup>4</sup> e uso de recursos. Ele agenda trabalhos e tarefas;
- *MapReduce* - Uma estrutura que ajuda os programas a fazerem cálculos paralelos com os dados. O trabalho a ser realizado obtém dados de entrada e os converte em um conjunto de dados que pode ser calculado em pares de valores-chave. Esse trabalho é dividido em tarefas que são realizadas em paralelo pelo cluster, permitindo se chegar ao resultado do trabalho em um tempo menor;
- *Hadoop Common* - Fornece bibliotecas Java<sup>5</sup> comuns que podem ser usadas em todos os módulos.

Associado ao Hadoop, os bancos de dados NoSQL são utilizados para atender as demandas das aplicações modernas da TIC, em especial as de *Analytics*. Os Banco de dados NoSQL são bancos de dados criados para fins específicos e que possuem esquemas flexíveis, segundo a AWS-Amazon (SD) os principais tipos são:

- Baseado em Coluna (*Column Stores*) - é otimizado para recuperação rápida de colunas de dados, normalmente em aplicativos analíticos. O armazenamento orientado a colunas para tabelas do banco de dados é um fator importante para a performance de consulta analítica, pois ele reduz drasticamente os requisitos de entrada(-gravação) e saída(leitura) nos discos. Principais produtos: Hbase, Cassandra, Hypertable, Accumulo, Amazon SimpleDB, Cloudata, Cloudera, SciDB, HPCC, Stratosphere;
- Baseado em Documentos (*Document Stores*) - projetado para armazenar e consultar dados como documento. Facilitam para que os desenvolvedores armazenem e consultem dados usando o mesmo formato de modelo de documento

<sup>3</sup> *Low-end* são equipamentos de baixo custo.

<sup>4</sup> *Cluster* em TI é a conexão de diversos equipamentos com o objetivo de aumentar desempenho e a tolerância a falhas.

<sup>5</sup> Java é uma linguagem de programação orientada a objeto desenvolvida por James Gosling, juntamente com outros colaboradores, no início da década de 1990, na empresa Sun Microsystems, hoje, Oracle.

<sup>2</sup> *Open Source* - termo criado pela *Open Source Initiative* para designar um programa de computador (*software*), onde é livre a consulta, avaliação e modificações sem incorrer em problemas de direito autoral, o que promove um modelo colaborativo de desenvolvimento de programa de computador.

que usam no código do aplicativo. Principais produtos: MongoDB, CouchDB, BigCouch, RavenDB, Clusterpoint Server, ThruDB, TerraStore, RaptorDB, JasDB, SisoDB, SDB, SchemaFreeDB, djondb;

- Baseado em Grafos (*Graph-Based Stores*) - especializados em armazenar e recuperar informações utilizando estruturas de vértices e arestas, evidenciam relacionamentos entre objetos, sejam eles pessoas, ativos de rede ou mesmo cidades. Principais produtos: Neo4J, Infinite Graph, Sones, InfoGrid, HyperGraphDB, DEX, Trinity, AllegroGraph, BrightStarDB, BigData, Meronymy, OpenLink Virtuoso, VertexDB, FlockDB;
- Baseado em Chave-Valor (*Key-Value Stores*) – são altamente particionáveis e permitem escalabilidade horizontal em escalas que outros tipos de banco de dados não conseguem alcançar. Principais produtos: Dynamo, Azure Table Storage, Couchbase Server, Riak, Redis, LevelDB, Chordless, GenieDB, Scalaris, Tokyo Cabinet/Tyrant, GT.M, Scalien, Berkeley DB, Voldemort, Dynamite, KAI, MemcacheDB, Faircom C-Tree, HamsterDB, STSdb, Tarantool/Box, Maxtable, Pin caster, RaptorDB, TIBCO Active Spaces, allegro-C, nessDB, HyperDex, Mnesia, LightCloud, Hibari, BangDB.

Para a AWS-AMAZON(SD), os bancos de dados NoSQL são ótimas ferramentas para muitas aplicações modernas para WEB, para jogos e para ambientes moveis, pois oferecem:

- Flexibilidade: os bancos de dados NoSQL geralmente fornecem esquemas flexíveis que permitem um desenvolvimento mais rápido e mais iterativo. O modelo de dados flexível torna os bancos de dados NoSQL ideais para dados semiestruturados e não estruturados;
- Escalabilidade: os bancos de dados NoSQL geralmente são projetados para escalar usando clusters distribuídos de hardware em vez de ampliá-los, adicionando servidores caros e robustos. Alguns provedores de nuvem lidam com essas operações nos bastidores como um serviço totalmente gerenciado;
- Alto desempenho: o banco de dados NoSQL é otimizado para modelos de dados específicos (como documento, valor-chave e gráfico) e padrões de acesso que permitem maior desempenho do que tentar realizar uma

funcionalidade semelhante com bancos de dados relacionais;

- Altamente funcional: os bancos de dados NoSQL fornecem APIs<sup>6</sup> e tipos de dados altamente funcionais criados especificamente para cada um de seus respectivos modelos de dados.

Os princípios de Hadoop e NoSQL exigem ambientes de hardware multiprocessados de alto poder de processamento e que fazem uso de grandes quantidades de memórias, variando, em muito, essas necessidades de acordo com o tipo de uso que se está fazendo em um determinado momento. Esse ambiente volátil, faz com que as organizações tenham dificuldades em manter seus parques de equipamentos de forma adequada. Por isso, cada vez mais, as aplicações de *Big Data Analytics* ocorrem em ambientes de *Datacenters*<sup>7</sup> e *Cloud Computing*<sup>8</sup>. Isso se dá pelo poder de escalonamento que esses oferecem, pois em um momento de maior uso, automaticamente podem aumentar o poder de processamento e findado esse momento, reduzir aos padrões originais. Sendo que a organização paga somente pelo que usou, permitindo uma melhor equalização dos custos da TICs.

Apesar do Hadoop e do NoSQL serem os principais itens de software em um ambiente *Big Data Analytics*, a cada dia surgem novas ferramentas *Open Source* para ajudar a controlar e tirar o máximo proveito desse ambiente. No entanto, o ser humano ainda é o principal fator nesse processo e o profissional ligado a esse ambiente recebe o nome de *Data Scientist* (cientista de dados).

<sup>6</sup> *Application Programming Interface* é um conjunto de ferramentas, definições e protocolos para criar e integrar softwares. Com ela, uma solução ou serviço pode se comunicar de forma simples com outras soluções ou serviços sem saber como essas foram feitas as suas implementações.

<sup>7</sup> *Datacenter* é um ambiente computacional que agrega toda a infraestrutura necessária para que os sistemas de organização sejam processados, envolvendo armazenamento, acesso a Internet e as aplicações. Esse ambiente pode ser privado, ou seja, pertencente a uma única organização, ou pode ser público, uma empresa oferece os seus recursos a diversas outras empresas, cobrando apenas pelos recursos efetivamente utilizados.

<sup>8</sup> *Cloud Computing* ou Computação em Nuvem refere-se ao armazenamento e processamento de dados fora dos seus locais de origem, sendo acessado através do uso da Internet. Como o usuário não sabe exatamente onde esses dados estão armazenados e sendo processados, criou-se a ideia de que estão na nuvem, ou seja, em qualquer lugar. A nuvem pode ser: privada, onde os recursos de ambiente computacional são da própria organização e só ela tem acesso; pública, onde os recursos do ambiente computacional são armazenados *Datacenters* públicos; Híbrido, há uma mistura entre o uso da nuvem privada e da nuvem pública.

## DATA SCIENTIST.

Com o *Big Data Analytics*, surgiu o termo *Data Scientist* ou Cientista de dados para designar aquele profissional que dá um bom uso ao Data Lake sobre o qual se assenta, permitindo novas percepções e *insights* que promovam a inovação e a melhor compreensão do ambiente que cerca as organizações. Ele é um profissional especializado em dados analíticos e que faz uso de suas habilidades em tecnologias e em ciências sociais aplicadas para encontrar tendências e gerenciar dados.

Segundo Amaral (2016, p.13), esse termo descrevia o profissional com conhecimento técnico vertical em estatística, NoSQL, *cloud Computing*, mineração de dados, álgebra relacional, MapReduce, virtualização, entre outros.

No momento atual, além das expertises apresentadas o cientista de dados também necessita possuir habilidades que permitam realizar uma série de atividades, para SAS(SD), são elas:

- Coletar grandes quantidades de dados não-estruturados e transformá-los em um formato mais utilizável;
- Resolver problemas de negócios usando técnicas orientadas por dados;
- Trabalhar com uma variedade de linguagens de programação, incluindo SAS, R e Python;
- Ter uma sólida compreensão de estatística, incluindo testes e distribuições;
- Estar sempre atualizado sobre técnicas analíticas, como machine learning, deep learning e análise de texto;
- Comunicar e colaborar tanto com TI quanto com a gerência;
- Procurar por ordem e padrões nos dados, além de identificar tendências que podem ajudar no resultado financeiro de uma empresa.

Encontrar todas essas características em um único profissional é muito difícil, tornando essa carreira aberta a multidisciplinaridade.

Com a evolução dos *softwares* voltados a *Big Data Analytics*, a parte técnica fica cada vez mais reduzida a um seleto grupo de profissionais altamente qualificados, abrindo espaços para especialistas das mais diversas áreas atuarem na parte de análise

gerados para visualização, mas que requerem uma análise mais aprofundadas.

## CIENTISTA SOCIAL.

Segundo o Projeto Político Pedagógico, do Curso de Ciências Sociais da UNESP de Araraquara, espera-se que o profissional por ele formado seja capaz de:

- elaborar, supervisionar, coordenar, planejar, programar, implantar, controlar, dirigir, executar, analisar ou avaliar estudos, trabalhos, pesquisas, planos programas e projetos atinentes à realidade social;
- ensinar sociologia geral ou especial nos estabelecimentos de ensino, desde que cumpridas as exigências legais;
- assessorar e prestar consultoria a empresas, órgãos da administração pública direta ou indireta, entidades e associações, relativamente à realidade social;
- participar da elaboração, supervisão, orientação, coordenação, planejamento, programação, implantação, direção, controle, execução, análise ou avaliação de qualquer estudo, trabalho, pesquisa, plano, programa ou projeto global, regional ou setorial, atinente à realidade social (Decreto nº 89.531, de 05 de abril de 1984).

Em uma sociedade complexa, multicultural e de mudanças constantes o papel do Cientista Social é essencial para ajudar a compreendê-la. Para Guiddens(1991, p. 12), a sociedade atual apresenta descontinuidades entre as instituições modernas e as tradicionais, ele identifica as características envolvidas:

Diversas características estão envolvidas. Uma é o *ritmo de mudança* nítido que a era da modernidade põe em movimento. As civilizações tradicionais podem ter sido consideravelmente mais dinâmicas que outros sistemas pré-modernos, mas a **rapidez da mudança** em condições de modernidade é extrema. Se isto é talvez mais óbvio no que toca à **tecnologia**, permeia

também todas as outras esferas. Uma segunda descontinuidade é o *escopo da mudança*. Conforme diferentes áreas do globo são postas em **interconexão**, ondas de transformação social penetram através de

virtualmente toda a superfície da Terra. (Negritos por nossa conta).

Nesse contexto, é o cientista social que consegue perceber os caminhos que a sociedade toma, os seus porquês e possíveis cenários futuros. Cabe ao Cientista Social estudar, compreender e defender os direitos humanos que reafirmam a nossa identidade humana.

Para as organizações, ambientes de incertezas podem ser muito prejudiciais aos negócios. Reduzir as incertezas pode significar a diferença entre o crescimento ou o fim de uma organização. Ter um profissional que a ajude a navegar por esse mar revolto torna-se cada vez mais relevante, o Cientista Social pode ser esse profissional, pois com suas expertises ele pode auxiliá-la a compreender e avaliar os desafios que advém dessa sociedade. Neste contexto, para Braga (2014, p. 142), o domínio da utilização de técnicas de pesquisa é o principal atrativo dos Cientistas Sociais para o mercado de trabalho.

Neste ambiente de alta complexidade aliada a maximização do volume de dados advindo a partir das revoluções das TICs modifica-se a própria forma de trabalho do profissional de Ciências Sociais. Para Ian et al. (2017, p.1):

O mundo tem mudado para os cientistas sociais. O Big Data está gerando todo um novo campo de pesquisa da ciência de dados. Esse mundo é dominado por cientistas da computação que geraram novas formas de criar e coletar dados, desenvolveram novas técnicas analíticas e estatísticas e acabaram fornecendo novas formas de visualizar e apresentar informações. Estas novas fontes de dados e técnicas têm o potencial de transformar a forma como a ciência social é feita. (tradução nossa).

Se antes, o cientista social, debruçava-se sobre livros, a compra e leitura de jornais e revistas, entrevistas pessoais, hoje tudo isso transformou-se. A maioria dos jornais e revistas impressos tem suas vendas cada vez mais reduzidas e as que sobrevivem são cada vez mais disponíveis *online*. O mesmo ocorre com os periódicos científicos. Livros cada vez mais são disponíveis em forma de *e-book* para serem lidos nas mais variadas formas de telas, inclusive afetando de forma negativa o mercado de livrarias físicas, enquanto o mercado de vendas *online* faz surgir novos *players* de grande alcance e poder econômico, como

a Amazon. As pesquisas, cada vez mais ocorrem em ambientes de *surveys online*. Essa realidade pode ser constatada nas palavras de Witte (2012, p. 56):

Para muitos cientistas sociais, hoje, visitas regulares às estantes da biblioteca, para localizar volumes empoeirados ou pesquisar nos arquivos e fichas de registro bibliográfico, são coisas do passado. Por outro lado, a busca e acesso, online a bases de dados digitais e a versões em texto integral de literatura científica são atividades cotidianas. Esta transformação foi tão completa e as novas práticas digitais tão ubíquas, que a magnitude dessa mudança é, muitas vezes, subestimada.

Esse imenso *Data Lake* produz uma variedade de dados jamais observadas pelo ser humano, informações advindas de redes sociais tais como Facebook, Instagram, Whatsapp, Twitter ou por sistemas de geolocalização como o Waze produzem dados que mostram os anseios e cotidiano de bilhões de pessoas e que se bem organizados e analisado proporcionam ao Cientista Social possibilidades de *insights* inovadores em pesquisas sociais. A partir dele pode-se compreender como comportamentos e hábitos sociais modificam-se ou como os caminhos políticos se transformam ao serem transportados das discussões locais para o ambiente de WEB e das redes sociais.

Observa-se nos tempos recentes os usos desses meios para disseminação das chamadas *Fakes News*, para comunicações governamentais e para debates acalorados sobre política e eleições tornou-se lugar comum, gerando amplo campo para discussões e pesquisas para os profissionais das Ciências Sociais.

Ao analisar resultados do *Big Data*, o cientista social poderá fazer inferências de como discussões em grupos de redes sociais e a proliferação de *Fake News* afetam e porque afetam eleições, como mudam os relacionamentos sociais, de amizade e mesmo amorosos, entender como a sociedade está transformando do mundo “físico” para o mundo “virtual”. Para Grimmer (2015, p. 80), o *Big Data* é a promessa de que nesses dados estão as respostas para questões fundamentais para os negócios, para os governos e para as Ciências Sociais.

Ainda aprimora e aprofunda a forma como o Cientista Social pode acompanhar e observar o crescentes desafios e desrespeitos pelo quais passam os direitos humanos, como afirma Schulz (2018, p.1):

A Declaração Universal dos Direitos Humanos chega aos 70 anos agora em 2018. No que deveria ser um período de celebração, assistimos, no entanto, a um desfile de constantes violações abjetas e brutais desses mesmos direitos que deveriam reafirmar a nossa identidade humana. Em vez disso, ameaça-nos a desumanização.

Nesse ambiente multifacetado, o profissional de Ciências Sociais e das Ciências Sociais Aplicadas se põe em confronto com outras profissões que disputam pelo mesmo espaço. Um dos maiores desafios que o profissional de Ciências Sociais enfrenta diz respeito a falta de barreiras que delimitem as funções das diversas profissões, para Bonelli (1993, p. 43)

As ciências sociais, como qualquer outra profissão, enfrentam competição direta com seus “vizinhos”. Em termos concretos, disputam-se objetos de estudo, vagas no mercado, formas específicas de abordar realidades que qualificam mais a profissionais de uma área que de outra. Disputa-se também a regulamentação que determine o monopólio do exercício das atividades profissionais.

Essa situação de disputa pelas oportunidades oferecidas pelo uso de *Big Data Analytics* repete-se frente aos chamados *Data Scientist*, como pode-se inferir a partir das palavras Grimmer (2015, p. 80)

Muitas das reivindicações mais turbulentas vêm de campos computacionais, que têm pouco experiência com a dificuldade da investigação científica social. Como cientistas sociais, podemos nos assegurar de que sabemos melhor. Nossa ampla experiência com dados observacionais significa que sabemos que grandes conjuntos de dados sozinhos são insuficientes para resolver os problemas mais prementes da sociedade. Nós até podemos ter de ministrar cursos sobre como seleção, medição erro e outras fontes de preconceito devem nos tornar céticos quanto a ampla gama de problemas. Esta afirmação é verdadeira; “Big data” sozinho é insuficiente para resolver os problemas mais urgentes da sociedade - mas certamente auxilia muito. (Tradução própria).

A capacidade de análise e reflexão, inerente a formação dos profissionais em Ciências Sociais, é fundamental para tornar os dados obtidos pelo *Big Data Analytics* de relevância para as organizações, governos e sociedade, apontando tendências, compreendendo situações, gerando novas visões sobre determinados ambientes em determinadas situações

e gerar novas teorias que sem acesso a essa grande quantidade de dados não haveria como serem desenvolvidas. Essas características tornam o profissionais das Ciências Sociais extremamente atraente para as mais diversas oportunidades de ocupações de atuação com *Big Data Analytics*, bem como um vasto campo para as pesquisas em Ciências Sociais.

O que irá garantir ao profissional de Ciências Sociais disputar e ampliar o seu mercado de trabalho é a sua capacidade de adquirir novas expertises. Essas novas aquisições passam por compreender os conceitos de *Big Data Analytics* e incorporá-los aos seus saberes, permitindo uma melhor análise dos ambientes onde ocorrem as relações sociais modernas.

## CONSIDERAÇÕES FINAIS.

O *Big Data* proporciona a humanidade acesso a um volume de informações jamais visto anteriormente, mantendo um crescimento exponencial que, até o momento, parece não se esgotar. Esse novo ambiente de grande volume dados, originados das mais variadas fontes e acessados a altíssima velocidade exigiu que novas tecnologias surgissem e novos paradigmas na área da TIC fossem desenvolvidos dando origem a termos como Hadoop, NoSQL, IoT, *Cloud Computing* e outros tantos, que cada vez mais fazem parte do dia a dia das organizações e do cidadão comum.

Esse enorme *Data Lake* permite que percepções e predições sejam feitas com um grau de precisão cada vez maior, melhorando a tomada de decisões, a criações de estratégias e a compreensão do que acontece e está por acontecer nas mais variadas áreas do saber.

Apesar de tecnicamente ainda ser um mundo voltado a especialista, todos os dias surgem novas ferramenta que automatizam e melhoram os processos de *Big Data Analytics*, principalmente as ferramentas voltadas para a visualização de dados. Essas ferramentas permitem que cada vez mais seja reservado ao ser humano a parte nobre do processo: pesquisar, analisar, criar e melhorar métodos e gerar novos *insights*. Isso resulta em oportunidades para que novos jogadores participem desse ambiente, fazendo que não seja mais exclusividade de um grupo seletivo de profissionais altamente especializado em técnicas computacionais, mas que envolvam cada vez mais profissionais que tenham a capacidade de compreender socialmente o



que esses dados estão dizendo, o que representam e o seus impactos na sociedade e para os negócios.

Esse é um campo que se abre as mais diversas profissões, em especial a aquelas que em seu bojo está a pesquisa e a análise. Dentro desse ambiente, os profissionais das Ciências Sociais e das Ciências Sociais Aplicadas já possuem suas formações os *skills* para pesquisa e para a análise, devendo buscar o que lhe falta para adentrar no Big Data Analytics. Isso lhe permitirá, fazer pesquisa e análises com grau de certeza maior, aumento a qualidade dos resultados obtidos.

Compreender esse ambiente é essencial a quem quer fazer pesquisa e análise sobre a sociedade, o momento em que vivemos e para onde vamos.

Caso o profissional de Ciências Sociais se modele a esse ambiente e aprenda a extrair mais e melhores resultados a partir do Big Data Analytics, um amplo campo de trabalho se desabrocha, sendo um profissional altamente requisitado nas mais variadas áreas, desde a pesquisa pura tradicional até o trabalho em grandes organizações transnacionais. Caso contrário, sofre o perigo de ter suas atividades sobrepujada por outras categorias profissionais aumentada, perdendo espaço em um mercado de trabalho cada vez mais concorrido, mais desregulamentado, precário e tecnológico.

Se por um lado, o *Big Data Analytics* gera novos desafios para os profissionais de Ciências Sociais e das Ciências Sociais Aplicadas, por outro, se abre um mundo novo de oportunidades, onde a sua capacidade de pesquisa, análise crítica e capacidade de inferência são essenciais.

## REFERÊNCIAS.

- Amaral, Fernando. **Introdução à Ciência de Dados: Mineração de Dados e Big Data.** Books. 2016.
- AWS-Amazon. **What is Hadoop?** Disponível online em <https://aws.amazon.com/pt/emr/details/hadoop/what-is-hadoop/> - Acesso em 19/04/2019. S/d.
- AWS-Amazon. **O que é NoSQL?** Disponível online em <https://aws.amazon.com/pt/emr/details/hadoop/what-is-hadoop/> - Acesso 17/04/2019. S/d.
- Bonelli, Maria da Glória. **Identidade Profissional e Mercado de Trabalho dos Cientistas Sociais:** As Ciências Sociais no Sistema da Profissões. Tese de Doutorado apresentado ao Departamento de Ciências Sociais do Instituto de Filosofia e Ciências Humanas da Universidade Estadual de Campinas. Campinas, 1993.
- Braga, Eugenio Carlos Ferreira. **Cientistas Sociais extra-universitários: Identidade Profissional no Mercado de Pesquisa.** Estudos de Sociologia. Araraquara. V.14, n. 26, p. 141-167. 2009.
- Desjardins, Jeff. **What happens in a Internet Minute in 2019?** Disponível em: <https://www.visualcapitalist.com/what-happens-in-an-internet-minute-in-2019/> . Acesso em 17 de abril de 2019. S/d.
- Evans, Dave. **A Internet das coisas:** Como a próxima evolução da Internet está mudando tudo. Disponível on line em [https://www.cisco.com/c/dam/global/pt\\_br/assets/executives/pdf/internet\\_of\\_things\\_iot\\_ibsg\\_0411final.pdf](https://www.cisco.com/c/dam/global/pt_br/assets/executives/pdf/internet_of_things_iot_ibsg_0411final.pdf) . 2011. Acesso em 15/05/2019.
- Ian, Foster; Ghani,Rayid; Jarmin, Ron S.; Kreuter, Frauke; Lane, Julia. **Big Data and Social Science: A Pratical Guide to Methods and Tools.** CRC Press, Boca Raton. 2017.
- Grimmer, Justin. **We Are All Social Scientist Now: How Big Data, Machining Learning and Casual Inference Work Together.** American Social Science Association, 2015, Disponível on line em [https://stanford.edu/~jgrimmer/bd\\_2.pdf](https://stanford.edu/~jgrimmer/bd_2.pdf) doi:10.1017/S1049096514001784 , acesso em 18/05/2018.
- Guiddens, Anthony. **As conseqüências da Modernidade.** Editora Unesp. São Paulo. 1991.
- Ishwarappa, K; Anarudha, J. **A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology.** Elsevier. Science Direct. Procedia Computer Science. 48 315-324p. 2015. Disponível online em <https://reader.elsevier.com/reader/sd/pii/S1877050915006973?token=959B078DEFB9B71CBF2C9868C193D90D9645CF0CE1647057E6433C46C407181875BF30834642B702DE4C6FA333066818> acesso em 10/05/2019.
- Marr, Bernard. **What is a Data Lake?** A super-simple explanation for anyone. Forbes Midia. 2018. Disponível on line em <https://www.forbes.com/sites/bernardmarr/2018/08/27/what-is-a-data-lake-a-super-simple-explanation-for-anyone/#634a7f9676e0> - Acesso em 15/05/2019.
- McAfee, Andrew; Brynjolfsson, Erik. **Big Data: Management Revolution.** Harvard Business Review, 2012. Disponível online <http://tarjomefa.com/wp-content/uploads/2017/04/6539-English-TarjomeFa-1.pdf> – Acesso 17/04/2019.
- Praveen, Shagufta; Chandra, Umesh. **Influence of Structured, semi-structured, Unstructured data on various data models.** International Journal of Scientific & Engineering Research. V 8. 2017. Disponível online em <https://www.ijser.org/researchpaper/Influence-of->

Structured--Semi-Structured--Unstructured-data-on-various-data-models.pdf acesso em 10/05/2019.

ResearchGate. **Scientific Big Data and Digital Earth**. Disponível online em [https://www.researchgate.net/figure/Global-growth-trend-of-data-volume-2006-2020-based-on-The-digital-universe-in-2020\\_fig1\\_274233315](https://www.researchgate.net/figure/Global-growth-trend-of-data-volume-2006-2020-based-on-The-digital-universe-in-2020_fig1_274233315) - Acesso em 17/04/2019. S/d.

Russon, Philip. **Big Data Analytics**. TDWI Best Practices Report. Fourth Quarter 2011. Disponível on line em <https://vivomente.com/wp-content/uploads/2016/04/big-data-analytics-white-paper.pdf> - Acesso em 15/05/2019.

SAS. **Cientistas de dados: Quem são e por que você quer ser um**. Disponível on line em [https://www.sas.com/pt\\_br/insights/analytics/cientistas-de-dados.html](https://www.sas.com/pt_br/insights/analytics/cientistas-de-dados.html) - Acesso em 15/05/2019. S/d.

Schultz, Peter. **Ciência e os Direitos Humanos**. Jornal da Unicamp – Edição Web. Campinas, 2018. Disponível online em <https://www.unicamp.br/unicamp/ju/artigos/peter-schulz/ciencia-e-direitos-humanos> - Acesso em 20/05/2019.

Sicular, Svetlana. **Gartner's Big Data Definitions Consists of Three Parts Not to Be Confused with Three "V"s**. 2013. Disponível on-line em <https://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/#5e7ceb1942f6> acesso em 29/07/2019.

Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP - Campus Araraquara. **Projeto Político Pedagógico do Curso de Ciências Sociais**. Disponível online em <https://www.fclar.unesp.br/Home/Graduacao/PPP%20Ciencias%20Sociais.pdf> – Acesso em 15/05/2019. S/d.

WITTE, James C. **A Ciência Social Digitalizada: Avanços, oportunidades e desafios**. Sociologias. Porto Alegre, ano 14, nº 31, set/dez, 2012, p.52-92.