

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

GT-1 – Estudos Históricos e Epistemológicos da Ciência da Informação

REFLEXÕES CONCEITUAIS E ÉTICAS SOBRE BIG DATA: LIMITES E OPORTUNIDADES

Lisiane Braga Ferreira - (Universidade Federal do Rio Grande do Sul - UFRGS)

Moisés Rockembach - (Universidade Federal do Rio Grande do Sul - UFRGS)

Luciana Monteiro Krebs - (Universidade Federal do Rio Grande do Sul - UFRGS)

CONCEPTUAL AND ETHICAL REFLECTIONS ON BIG DATA: LIMITS AND OPPORTUNITIES

Modalidade da Apresentação: Pôster

Resumo: Este trabalho objetiva analisar a produção científica a respeito da temática *Big Data* em consonância com aspectos éticos que o fenômeno apresenta. Trata-se de uma pesquisa de natureza básica, metodologia descritiva e abordagem quanti-qualitativa, baseada em busca na plataforma *Web of Science* utilizando os tópicos "*big data e ethic**", e em revisão bibliográfica para análise dos dados. Os resultados obtidos mostram que, embora em tendência crescente, existem poucas pesquisas sobre *Big Data* e ética disponíveis na *Web of Science*, especialmente oriundas da Ciência da Informação. Evidenciam-se os principais agrupamentos temáticos dentro deste universo e os trabalhos mais citados pela comunidade científica. Também se observa que diferentes áreas do conhecimento, além da Ciência da Computação, vêm discutindo a temática. Foram elencados e discutidos os dilemas éticos levantados pelos autores mais citados na temática, como invasão de privacidade, diminuição das liberdades civis e aumento do controle estatal e corporativo. Também foi levantado o uso destes grandes volumes de dados como *commodities* e a necessidade de proteção da privacidade por meio de legislações específicas. Conclui-se que existe uma carência de produção sobre *Big Data* e ética, o que abre um campo profícuo para reflexões sobre os limites do uso e reuso do *Big Data* em nossa sociedade, a ser explorado pelos pesquisadores da Ciência da Informação.

Palavras-Chave: *Big Data*; Ética; Ciência da Informação.

Abstract: This paper analyzes the scientific production regarding the Big Data theme in consonance with the ethical aspects that the phenomenon presents. A quantitative-qualitative approach and descriptive methodology is used to collect, process and analyse the data from a search in the Web of Science platform, using the "big data and ethic *" topics. A bibliographic review was made for data analysis. The results show that, although in a growing trend, there is incipient research about Big Data and ethics available in the Web of Science, especially from Information Science. The main thematic clusters within this universe were shown, also the works most cited by the scientific community. It was also observed that different areas of knowledge, in addition to Computer Science, have been discussing the theme. The ethical dilemmas raised by the most cited authors on the issue were listed and

discussed, such as invasion of privacy, reduction of civil liberties and increased state and corporate control. It was also identified the use of these large volumes of data as commodities and the need to protect privacy through specific legislation. It is concluded that there is a lack of production on Big Data and ethics, which opens a productive field for reflections on the limits of the use and reuse of Big Data in our society, to be explored by researchers of Information Science.

Keywords: Big Data; Ethics; Information Science.

1 INTRODUÇÃO

Entre 1820 e 1840 houve uma explosão de números não ainda relacionada com o aumento de tecnologias digitais, mas se deu a partir de um "entusiasmo pelos números" e uma crescente coleta de dados populacionais que tornou o termo estatística indispensável. A sensação de dilúvio de dados sobre as pessoas tem uma longa história, onde os tipos de dados mudaram, mas a linhagem é clara. O início dessa avalanche de dados se deu pela urbanização das cidades, hoje o *Big Data* se potencializa pelas formas comerciais e algorítmicas. (HACKING, 1990; PORTER, 1986; BEER, 2016).

Muitos fatores impulsionam a produção e o compartilhamento de informações em rede, como a navegação móvel 3G, a aceleração da conexão em rede, bases de dados voltadas para a clusterização de dados (*Big Data/Hadoop*), desenvolvimento da inteligência artificial (IA) e outros recursos. Tecnologias como a Internet promoveram a comunicação horizontal, de cidadão para cidadão, e esta capacidade de comunicação em massa, não mediada pelos meios tradicionais de comunicação, também propiciou a produção de insumos para as grandes companhias de tecnologia e comunicação obterem lucros por meio de filtros e coleta de dados. Nossa interação constante gera o *Big Data*. (FRIEDMAN, 2006; CASTELLS, 2000).

Big Data é um volume muito grande de dados de vários tipos, coletados massivamente de fontes heterogêneas (sensores, telefones celulares, redes sociais etc.), mas também são as novas tecnologias (algoritmos/*software* e *hardware*) usadas para coletar, armazenar e processar esses dados com alta velocidade. O *Big Data* caracteriza-se por três aspectos principais, os 3V's: Velocidade, Volume e Variedade. (NERURKAR; WADEPHUL; WIEGERLING, 2016; ZIKOPOULOS *et al*, 2011). Ele vem sendo apresentado como solução para muitos problemas na Tecnologia da Informação que antes, ou não eram passíveis de solução, ou exigiam altos custos. Alguns cientistas de dados e pesquisadores acreditam que o *Big Data* promete um aumento na eficiência e produtividade, que tende a reduzir custos na indústria e

nos negócios, permitindo novos métodos e conhecimentos científicos, bem como melhor controle e regulação na vida pessoal e na governança.

A ética, segundo Bauman (1997), é o código moral que pretende ser o único conjunto de preceitos harmonicamente coerentes ao qual deve obediência toda pessoa moral. E os grandes temas da ética também são questões que vigoram na atualidade e apenas precisam ser vistos e tratados de maneira nova. As Tecnologias de Informação e Comunicação (Floridi, 2010), ao transformem o contexto em que as questões morais surgem, não só agregam novas dimensões interessantes a antigos problemas, mas nos levam a repensar, metodologicamente, os próprios fundamentos em que nossas posições éticas se baseiam. Assim, esta pesquisa visa investigar as publicações acadêmicas recentes sobre ética e *Big Data* buscando um panorama da produção científica atual para compreensão do que se está discutindo a respeito da geração massiva de dados e seus impactos nas práticas comerciais, sociais e científicas.

2 METODOLOGIA

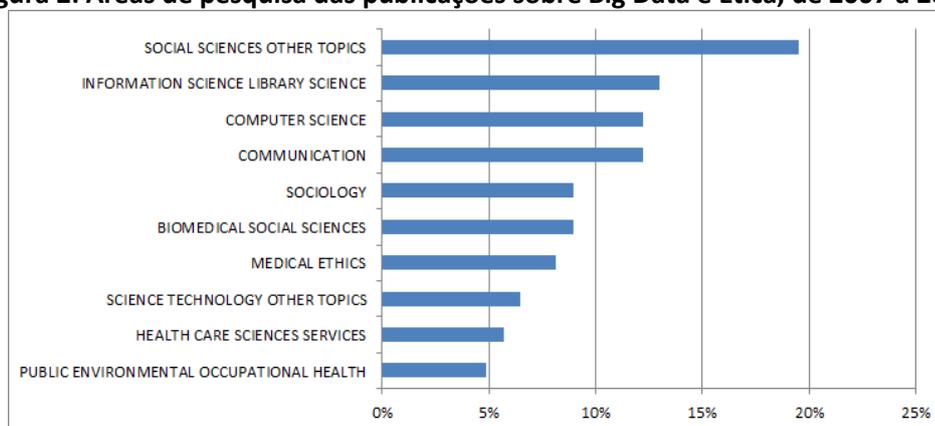
O estudo tem abordagem quanti-qualitativa e metodologia descritiva. Para obter os dados realizou-se uma busca, em 02 de agosto de 2017 na plataforma *Web of Science (WoS)*, pelo tópico “*big data*” juntamente com “ética” (termos em inglês e operador booleano, sendo a expressão de busca: “*big data*” AND *ethic**). Depois se filtrou o resultado por tipo de documento, selecionando apenas os artigos. O período escolhido foi de 2007 a 2016. Esta busca resultou em 131 artigos recuperados. Para geração da nuvem de palavras, coletaram-se os títulos dos trabalhos e aplicou-se uma *stopword list* para remoção de algumas categorias gramaticais (artigos, preposições, etc.). Foram então retiradas as palavras presentes na expressão de busca (nomeadamente “*big*”, “*data*”, “*ethical*” e “*ethics*”). Depois se contou a frequência das palavras e, por fim, criada a nuvem de palavras. Para a análise, selecionou-se de forma intencional bibliografia sobre *Big Data* e ética, a partir da pertinência ao tema proposto e campo científico.

3 ÉTICA E *BIG DATA*

A investigação caracterizou-se como pesquisa básica, que, segundo Vianna, Dutra e Frazzon (2016), que pesquisaram sobre a temática *Big Data*, é utilizada quando há escassez de produção relacionada aos termos e requer algumas investigações fundamentais para

(Figura 2). As áreas sociais, computação e medicina estão entre as que mais produzem sobre *Big Data* e ética.

Figura 2: Áreas de pesquisa das publicações sobre *Big Data* e Ética, de 2007 a 2016.



Fonte: Dados da pesquisa - 2017.

O *Big Data*, portanto, não é uma área de estudo exclusiva da Ciência da Computação pois, como afirmam Boyd e Crawford (2012), físicos, economistas, matemáticos, cientistas políticos, bio-informáticos, sociólogos e outros profissionais poderão fazer uso e reuso desta quantidade massiva de informações e das interações produzidas pelas pessoas no meio digital. Sobre os possíveis problemas éticos levantados, o *Big Data* pode ser visto como uma manifestação do *Big Brother*, trazendo consigo a invasão de privacidade, diminuição das liberdades civis e aumento do controle estatal e corporativo, gerando incerteza e medo quanto aos usos e manipulação dos dados (BOYD, CRAWFORD, 2012). Apresentamos abaixo os artigos mais citados, relacionando *Big Data* e ética.

Quadro 1: Artigos mais citados sobre *Big Data* e Ética, publicados entre 2007 e 2016.

Ranking	Título	Autores	Ano da publicação	Total de citações
1º	CRITICAL QUESTIONS FOR BIG DATA Provocations for a cultural, technological, and scholarly phenomenon	BOYD, D.; CRAWFORD, K.	2012	515
2º	Facebook as a Research Tool for the Social Sciences Opportunities, Challenges, Ethical Considerations, and Practical Guidelines	KOSINSKI, M. <i>et al.</i>	2015	26
3º	Nursing Needs Big Data and Big Data Needs Nursing	BRENNAN, P. F.; BAKKEN, S.	2015	24
4º	The commodification of patient opinion: the digital patient experience economy in the age of big data	LUPTON, D.	2014	24
5º	Critiquing Big Data: Politics, Ethics, Epistemology Special Section Introduction	CRAWFORD, K.; MILTNER, K.; GRAY, M. L.	2014	20
6º	A Survey of Big Data Research	FANG, H. <i>et al.</i>	2015	16
7º	Quantified sex: a critical analysis of sexual and reproductive self-tracking using apps	LUPTON, D.	2015	16
8º	Psycho-Informatics: Big Data shaping modern psychometrics	MARKOWETZ, A. <i>et al.</i>	2014	16
9º	The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care	COHEN, I. G. <i>et al.</i>	2014	15
10º	Acquisition, Analysis, and Sharing of Data in 2015 and Beyond: A Survey of the Landscape A Conference Report From the American Heart Association Data Summit 2015	ANTMAN, E. M. <i>et al.</i>	2015	14

Fonte: Dados da pesquisa - 2017.

Kosinski *et al.* (2015) abordam dilemas éticos em pesquisas sociais feitas no Facebook quanto a falta de agentes reguladores que realizem avaliações sem subestimar e nem superestimar as potenciais ameaças aos participantes. Essas questões desencorajam os cientistas sociais, abrindo a aplicação de pesquisas por cientistas da informática, que muitas vezes não se preocupam ou não estão familiarizados com as implicações éticas e sociais da Pesquisa de Assuntos Humanos (*Human Subjects Research - HSR*). Além disso, como apontado por Parisier (2012) a personalização e fragmentação do público nas redes sociais também podem gerar problemas éticos.

Lupton (2014) aborda uma economia de experiência emergente do paciente digital, na qual as experiências de doença e tratamentos médicos dos pacientes e suas opiniões sobre os cuidados de saúde e como o carregamento em plataformas *online* tornaram-se valorizadas e, em alguns contextos, monetizadas. Assim, o autor questiona o fato de que as pessoas que fornecem suas experiências não recebem compensação financeira. Por outro lado, as empresas com fins lucrativos que fornecem as plataformas para que os pacientes compartilhem suas experiências, lucram com os dados coletados. Estas plataformas tornaram-se *commodities* valiosas.

As preocupações com a privacidade tendem a influenciar a elaboração de legislações específicas, como a “*US Do Not Track Online Act*” em 2011 nos Estados Unidos (FANG *et al.*, 2015), permitindo que um usuário proíba a coleta ou uso de suas informações pessoais ou por exemplo com o “direito a ser esquecido” (*Right to be Forgotten*) na União Européia (BUNN, 2015).

Nerurkar, Waderphul e Wiegerling (2016) sugerem que o interesse generalizado na natureza do *Big Data* demonstra que há algumas dificuldades em dar uma definição concisa e informativa do conceito, especialmente com as Tecnologias de Informação e Comunicação (TIC). Estas evoluem rapidamente e, por vezes, estão sujeitas a avanços tecnológicos ou redefinições que transcendem conceitos de uso em um certo tempo. E a grande questão desloca-se de “quais são estes usos” para “quais são os limites destes usos” e “por que devemos pensar em uma ética para o *Big Data*?”.

As TIC, como aborda Floridi (2010), afetam a vida moral das pessoas (agentes) de inúmeras formas, que o autor representa em três linhas: informação como recurso, onde o agente pode aproveitar-se de algumas informações; informação como produto, onde o agente

gera outras informações e informação como alvo, afetando seu ambiente informativo. Segundo Floridi, este modelo simples serve como base de orientação inicial na multiplicidade de questões pertencentes à ética da informação. Conforme o autor, análises preditivas baseadas no Big Data (FLORIDI, 2012) também se configuram em potenciais problemas éticos.

Alguns temores éticos tendem a ser potencializados a partir do desenvolvimento da Internet das Coisas (do inglês *Internet of Things - IoT*), como Habib (2014) acrescenta, pois ao mesmo tempo em que a Internet trouxe comodidade para as vidas dos indivíduos, prometendo tornar a vida das pessoas muito mais fácil e conveniente, ela deixa aberta muitas “brechas” de insegurança. Esta característica propicia o rastreamento não autorizado e o monitoramento ilegal, afetando os relacionamentos de confiança, segurança e com implicações na privacidade individual.

4 CONSIDERAÇÕES FINAIS

O presente trabalho integra as análises desenvolvidas no projeto de pesquisa Epistemologia e ética do Big Data. Ao longo do trabalho verificamos uma carência de produção sobre *Big Data*, corroborando com Vianna, Dutra e Frazzon (2016), pois percebemos que a abordagem da temática encontra-se em estágio inicial nas pesquisas no campo da Ciência da Informação (representa 12% da produção analisada). Outra lacuna a ser investigada em trabalhos futuros está nos impactos éticos do uso e reuso do *Big Data*, reflexão que consideramos vital para a Ciência da Informação.

Percebemos que o avanço tecnológico, principalmente no âmbito da Internet, traz benefícios evidentes para a interação social e ao mesmo tempo cria problemas e dilemas éticos que por vezes possuem barreiras jurídicas ou legais em legislações vigentes, mas em alguns países (como o Brasil), as lacunas são mais visíveis. Estamos, talvez, no ápice de um momento “divisor de águas” – em que não podemos voltar ao que éramos antes da *web* nem seguir evoluindo sem delimitar preceitos regulatórios e limites éticos.

Há muitas discussões sobre o assunto, com argumentos contra e a favor. É perturbador, no entanto, que parece pouco viável a aplicação de um controle saudável para a democracia, para o direito ao acesso à informação, para o direito de expressão e para o direito à privacidade concomitantemente. Parece-nos sensato que pesquisas da Ciência da Informação agreguem esta temática para ampliar o debate e promover novas propostas que amenizem os impactos negativos da gestão e utilização da massa de dados. Uma visão crítica

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

sobre o *Big Data* nos levará a compreender as possibilidades de análise e construção de conhecimento, mas também as limitações éticas na coleta, análise e usos desses dados.

REFERÊNCIAS

BAUMAN, Z. **Ética pós-moderna**. São Paulo: Paulus, 1997.

BEER, David. How should we do the history of Big Data? **Big Data & Society**, 2016, v. 3, n. 1.

BOYD, D.; CRAWFORD, K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. **Information, communication & society**, v. 15, n. 5, p. 662-679, jun. 2012.

BUNN, A. The curious case of the right to be forgotten. **Computer Law & Security Review** v. 31, n. 3, jun. 2015.

CASTELLS, M. **Internet y la sociedade red**. [S/l]: UOC, 2000. Disponível em: <<http://www.uoc.edu/web/cat/articles/castells/castellsmain2.html>>. Acesso em: maio 2017.

FANG, H.; ZHANG, Z.; WANG, C. J.; DANESHMAND, M.; WANG, C.; WANG, H. A survey of big data research. **IEEE network**, v. 29, n. 5, p. 6-9, set/out. 2015.

FLORIDI, L. Big data and their epistemological challenge. **Philosophy & Technology**, v. 25, n. 4, 2012.

FLORIDI, L. **Information: A Very Short Introduction**. New York: Oxford University Press, 2010. 160 p.

FRIEDMAN, T. L. **Thank You for Being Late: An Optimist's Guide to Thriving in the Age of Accelerations**. What the Hell Happened in 2007. New York: Penguin Random House, 2016. p. 19-35.

HABIB, K. Ethical Aspects of the Internet of Things in eHealth. **International Review of Information Ethics (IRIE)**. v. 22, p. 83-91, dez. 2014.

HACKING, I. **The Taming of Chance**. Cambridge: Cambridge University Press, 1990.

KOSINSKI, M.; MATZ, S. C.; GOSLING, S. D.; POPOV, V.; STILLWELL, D. Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. **American Psychologist**, v. 70, n. 6, p. 543-556, set. 2015.

LUPTON, D. The commodification of patient opinion: the digital patient experience economy in the age of big data. **Sociology of health & illness**, v. 36, n. 6, p. 856–869, jul. 2014.

NERURKAR, M.; WADEPHUL, C.; WIEGERLING, K. Ethics of Big Data: Introduction. **International Review of Information Ethics (IRIE)**, v. 24, maio/2016.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

PARISER, Eli. **O filtro invisível**: o que a internet está escondendo de você. Rio de Janeiro: Zahar, 2012.

PORTER, T. M., **The Rise of Statistical Thinking 1820–1900**. Princeton, NJ: Princeton University Press, 1986.

VIANNA, W. B.; DUTRA, M. L.; FRAZZON, E. M. Big data e gestão da informação: modelagem do contexto decisional apoiado pela sistemografia. **Informação & Informação**. Londrina, v. 21, n. 1, p. 185-212, jan./abr. 2016.

ZIKOPOULOS, P.; Eaton, C.; DEROSS, D.; DEUTSCH, T.; LAPIS, G. **Understanding big data: Analytics for enterprise class hadoop and streaming data**. New York / Chicago / San Francisco: McGraw-Hill Osborne Media, 2011. Disponível em: <<http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=SA&subtype=WH&htmlfid=IML14296USEN>>. Acesso em: ago. 2017.