Confiamos nos dados? As implicações da datificação para o monitoramento social

In data we trust? The implications of datafication for social monitoring

JOSÉ VAN DIJCK[a]

Universiteit Utrecht. Utrecht, Países Baixos

RESUMO

Hoje há uma notável tolerância ao Big Brother e ao Big Business acessando rotineiramente as informações pessoais do cidadão, também conhecidas como Big Data. Parte da explicação para isso pode ser encontrada na gradual normalização da *datificação* como um novo paradigma na ciência e na sociedade. A datificação está se tornando um princípio central, não apenas entre os adeptos da tecnologia, mas também entre os acadêmicos. Este artigo desconstrói as bases ideológicas da datificação, argumentando que ela baseia-se em reivindicações ontológicas e epistemológicas problemáticas. A ideologia do *dataísmo* mostra características de crença generalizada na quantificação objetiva do comportamento humano, por meio das tecnologias de mídia on-line.

Palavras-chave: Big Data, datificação, vigilância, mídia social, metadados

ABSTRACT

Today there is a remarkable tolerance for Big Brother and Big Business routinely accessing citizens' personal information also known as Big Data. Part of the explanation for this may be found in the gradual normalization of *datafication* as a new paradigm in science and society. Datafication is becoming a leading principle, not just amongst techno-adepts, but also amongst scholars. This article deconstructs the ideological grounds of datafication arguing that in many respects datafication is rooted in problematic ontological and epistemological claims. The ideology of *dataism* shows characteristics of a widespread belief in the objective quantification of human behavior through online media technologies.

Keywords: Big Data, datafication, surveillance, social media, metadata

[a] Professora emérita da Universiteit Utrecht e presidente da Academia Real de Artes e Ciências dos Países Baixos. Autora de vários artigos e livros, entre outros, *The culture of connectivity: a critical history of social media*, publicado em 2013 pela Oxford University Press. E-mail: jose.van.dijck@knaw.nl

Confiamos nos dados? As implicações da datificação para o monitoramente social

CONFIAMOS EM DEUS. Todos os outros, nós monitoramos é uma famosa máxima empregada, pela primeira vez, pelo Serviço de Inteligência da Marinha dos Estados Unidos e depois utilizada pela Agência Nacional de Segurança (National Security Agency – NSA). Quando Edward Snowden, em 10 de junho de 2013, revelou ser o responsável pelos vazamentos que expuseram as práticas rotineiras de vigilância da NSA às novas mídias, descreveu em detalhe a *arquitetura da opressão* que permitiu a ele e muitos outros funcionários da NSA espiar os metadados de três bilhões de chamadas telefônicas e interações registradas pelo Facebook, Google, Apple e outras companhias de tecnologia. Numa entrevista em vídeo, o ex-analista da CIA disse que não poderia mais viver com a ampla invasão de privacidade e violações legais que tinha de realizar, atuando na comunidade de inteligência. Ele também queria tornar as pessoas conscientes do fato de que muitos agentes tinham acesso integral a todos os tipos de dados de comunicação, com o desejo de desencadear um debate público.

As revelações de Snowden eram, de fato, um alerta para os cidadãos que tinham aceitado gradualmente o *compartilhamento* de informações pessoais – tudo, desde o estado civil a resfriados, até os hábitos alimentares e música favorita – por meio de sites de redes sociais ou aplicativos, como a nova norma (van Dijck, 2013b). As plataformas proprietárias costumeiramente compartilham os metadados agregados de seus usuários com terceiros, com o propósito de marketing personalizado em troca de serviços gratuitos. Muitas pessoas não tinham percebido, até os vazamentos de Snowden, que as corporações das redes sociais também – desejosa ou relutantemente – compartilham suas informações com agências de inteligência. Quando Barack Obama defendeu suas políticas administrativas de vigilância em massa, dizendo que não havia "conteúdo, apenas metadados" envolvidos no sistema Prism, acrescentou que os cidadãos não poderiam esperar cem por cento de segurança, cem por cento de privacidade, e nenhum inconveniente. A explicação do presidente ecoava o argumento das companhias de mídia sociais de que os usuários lhes forneciam parte de sua privacidade em troca de convenientes plataformas de serviços gratuitos. Em outras palavras, os metadados parecem ter se tornado a *moeda corrente* para os cidadãos pagarem por seus serviços de comunicação e segurança – um desconfortável equilíbrio se instalara na zona de conforto da maioria das pessoas.

O que explica essa notável tolerância ao Big Brother e ao Big Business acessando rotineiramente as informações pessoais do cidadão, também conhecidas como Big Data? Parte da explicação pode ser encontrada na gradual normalização da *datificação* (*datafication*) como um novo paradigma na ciência e na

sociedade. A datificação, de acordo com Mayer-Schoenberger e Cukier (2013), é a transformação da ação social em dados on-line quantificados, permitindo assim monitoramento em tempo real e análise preditiva. As empresas e as agências governamentais exploram as pilhas exponencialmente crescentes de metadados coletados a partir da mídia social e plataformas de comunicação, tais como Facebook, Twitter, LinkedIn, Tumblr, iTunes, Skype, YouTube, e serviços gratuitos de e-mail, como o Gmail e o Hotmail, para rastrear informações sobre o comportamento humano: "Podemos agora coletar informações que não podíamos antes, seja sobre os relacionamentos revelados por chamados telefônicas ou sentimentos mostrados em *tweets*" (Mayer-Schoenberger; Cukier, 2013: 30). A datificação, como um legítimo meio para *acessar*, *entender* e *monitorar* o comportamento das pessoas está se tornando um princípio central, não apenas entre os adeptos da tecnologia, mas também entre os acadêmicos que a veem como uma revolucionária oportunidade de pesquisa para investigar o comportamento humano.

Neste artigo, pretendo desconstruir as bases ideológicas da datificação como definida por Mayer-Schoenberger e Cukier (2013) e ecoada por muitos proponentes desse novo paradigma científico. Argumentarei que a datificação é, em muitos aspectos, baseada em reivindicações ontológicas e epistemológicas problemáticas. Por mais convincentes que sejam alguns exemplos de Big Data em pesquisa, a ideologia do *dataísmo* (*dataism*) mostra características de *crença* generalizada na quantificação objetiva e o potencial monitoramento de todos os tipos de comportamento humano e de sociabilidade, por meio de tecnologias de mídia on-line. Além disso, o dataísmo envolve também a *confiança* nos agentes (institucionais) que coletam, interpretam e compartilham os (meta)dados extraídos da mídia social, das plataformas da internet e outras tecnologias de comunicação.

As noções de *confiança* e *crença* são particularmente relevantes quando se trata de entender a *vigilância de dados* (*dataveillance*): uma forma de contínua vigilância a partir do uso de (meta)dados (Raley, 2013). Como os documentos de Snowden deixaram claro, as pessoas têm fé nas instituições que lidam com seus (meta)dados, presumindo que elas seguem um conjunto de regras estabelecidas por agentes públicos responsáveis. Porém, como os jornalistas descobriram, a NSA constantemente desobedecia as decisões judiciais sobre o uso de dados, assim como as corporações estão constantemente testando os limites legais sobre a invasão de privacidade[1]. De modo mais profundo, o caso Snowden alertou mais ainda as pessoas para as práticas inter-relacionadas da inteligência do governo, empresas e da academia na adaptação das premissas ideológicas do dataísmo. Assim, precisamos olhar para a credibilidade de todo o ecossistema de mídia

[1] O grupo de defesa Consumer Watchdog, em maio de 2013, deu início a uma ação contra o Google, alegando que ele abre, lê e toma posse ilegalmente do conteúdo privado das mensagens de e-mail das pessoas.

conectiva. Quais são os distintos papéis do governo, corporações e academia em lidar com nossos dados? E qual tipo de atitude crítica é necessária frente a esse complexo sistema de fluxos de informação on-line?

DATIFICAÇÃO E "MINERAÇÃO DA VIDA" COMO UM NOVO PARADIGMA CIENTÍFICO

Ao longo da última década, a datificação tem se tornado um novo paradigma aceito para entender a sociabilidade e o comportamento social. Com o advento da Web 2.0 e os seus crescentes sites de redes sociais, muitos aspectos da vida social foram codificados, e o que nunca tinha sido quantificado – amizades, interesses, conversações casuais, buscas por informação, expressão de gostos, respostas emocionais, e assim por diante. À medida que as empresas de tecnologia começaram a se especializar em um ou vários aspectos da comunicação on-line, elas convenceram muitas pessoas a transferir parte de suas interações sociais para os ambientes da web. O Facebook tornou atividades sociais como *amizade* e *expressão de gostos* em relações algorítmicas (Bucher, 2012; Helmond; Gerlitz, 2013); o Twitter popularizou as personalidades on-line das pessoas e promoveu a ideia de criar *seguidores* e as funções de *retweet* (Kwak et al., 2010); o LinkedIn traduziu as redes profissionais de empregados e candidatos a vagas em interfaces digitais (van Dijck, 2013a); e o YouTube datificou a troca casual de conteúdo audiovisual (Ding et al., 2011). As interações sociais quantificadas tornaram-se, subsequentemente, acessíveis a terceiros: as próprias empresas, agências governamentais ou outras plataformas. A transformação digital da sociabilidade produziu uma indústria na qual seus progressos estão baseados no valor dos metadados – relatórios automatizados de quem se comunicou com quem, a partir de qual local e por quanto tempo. Os metadados – não há muito tempo, considerados subprodutos inúteis de serviços mediados por plataformas – gradualmente têm se tornado recursos valiosos que podem ser, ostensivamente, explorados, enriquecidos e reelaborados em produtos preciosos.

A perspectiva da indústria voltada aos dados não ressoa apenas nas auspiciosas metáforas de corrida do ouro dos empreendedores, mas também nas defesas dos investigadores que saúdam o Big Data como o santo graal do conhecimento comportamental. Os dados e metadados coletados do Google, Facebook e Twitter são, geralmente, considerados *impressões* ou *sintomas* dos comportamentos ou humores reais das pessoas, sendo as plataformas apresentadas como simples facilitadoras neutras. O Twitter supostamente permite a datificação das emoções, pensamentos e sentimentos viscerais das pessoas, já que a plataforma registra reações *espontâneas*; os usuários deixam marcas inconscientemente, de modo que

os dados podem ser "coletados passivamente sem muito esforço ou até mesmo consciência por parte daqueles que estão sendo gravados" (Mayer-Schoenberger; Cukier, 2013: 101). Os analistas com frequência descrevem a mediação em larga escala de *tweets* como similar ao uso de um termômetro para medir sintomas de febre nas multidões que reagem a acontecimentos sociais ou naturais – uma suposição baseada na ideia que o de que o tráfego social on-line flui em canais tecnológicos neutros. Nessa linha de raciocínio, nem a mediação tecnológica do Twitter por *hashtags*, *retweets*, algoritmos e protocolos, nem seu modelo de negócios parece relevante (Gillespie, 2010).

Os pesquisadores que endossam o paradigma da datificação tendem a ecoar essas afirmações a respeito da natureza da mídia social como pegadas naturais e as plataformas como facilitadores neutros. Os cientistas da informação têm chamado o Twitter de *sensor* de eventos em tempo real, ao processarem os *tweets* das pessoas sobre terremotos ou outros desastres (Sakaki; Okazaki; Matsuo, 2010). O Twitter também foi denominado um "detector de sentimento" das predileções políticas das pessoas (O'Connor et al., 2010) e um ferramenta que ajuda a entender as "dinâmicas de sentimento" a partir da análise de reações dos usuários do Twitter a um fragmento de vídeo específico (Diakopoulos; Shamma, 2010; Bollen; Mao; Pepe, 2011). A avaliação de grandes conjuntos de dados de plataformas de mídia social é cada vez mais apresentada como o mais escrupuloso e compreensivo método para medir a interação cotidiana, superior à amostragem (N=todos) e mais confiável do que as entrevistas ou levantamentos. Grandes quantidades de dados *desordenados* substituem as pequenas quantidades de dados amostrados e, como os defensores afirmam, o grande número de conjuntos de dados compensa a desordem dos mesmos. Alguns cientistas da informação defendem que o Twitter é, na verdade, uma enorme ferramenta de levantamento de opiniões em tempo real, pronta para tornar-se "um substituto e suplemento para os tradicionais levantamentos" (O'Connor et al., 2010). Existem importantes paralelos entre os levantamentos e os dados do Twitter, e as correlações encontradas nos resultados do Twitter são obviamente significativas. No entanto, as ressalvas sobre a alegada representatividade dessa rede social e os seus vieses (tecnológicos e comerciais) são parcamente abordados[2].

Os entusiastas da datificação também assumem com frequência uma autoevidente relação entre os dados e as pessoas, interpretando subsequentemente dados agregados para predizer comportamentos individuais. Por exemplo, Quercia et al. (2011) analisaram as relações entre a personalidade e os diferentes tipos de usuários do Twitter, descobrindo que os populares e influentes era também *imaginativos* e *organizados*. A partir desses padrões, eles especulam

[2] Algumas observações sobre a suposta representatividade do Twitter e seus inerentes vieses: a base de seus usuários não corresponde aos dados demográficos do público em geral. Um relatório do Pew Internet & American Life Project, publicado em fevereiro de 2012, descobriu que apenas 15% dos adultos on-line usam o Twitter e apenas 8% utilizam-no diariamente (Smith; Brenner, 2012). Além disso, o Twitter utiliza vários algoritmos que favorecem usuários influentes e permitem a manipulação das mensagens de *tweets*, tanto pela própria plataforma quanto por certas combinações de grupos de usuários (ver Cha et al., 2010). Os dados do Twitter são muitas vezes vistos como equivalentes a resultados de pesquisas, apesar das explícitas inadequações do valor representacional da ferramenta. Por exemplo, o Twitter Political Index ou Twindex, lançado em janeiro de 2012, rastreia os *tweets* que façam menção a um candidato. Twindex, uma parceria entre o motor de busca Topsy do Twitter e um par bipartidário de cientistas políticos, buscando medir "as mutações nos ânimos públicos, bem como consolidá-lo como uma plataforma para o debate cívico". Ver: <https://election.twitter.com/>.

quais usuários podem ter sucesso recomendando produtos ou ajudando a impulsionar estratégias de marketing. Num perspectiva similar, um recente estudo de Kosinski e colaboradores (2013) mostra como atributos e características particulares são previsíveis a partir dos registros digitais do comportamento humano, neste caso, os *likes* do Facebook eram usados para "automática e acuradamente predizer uma série de atributos pessoais altamente sensíveis, incluindo: orientação sexual, etnicidade, valores religiosos e políticos, traços de personalidade, inteligência, felicidade, uso de substâncias viciantes, separação dos pais, idade e gênero" (Kosinski; Stillwell; Graepel, 2013: 1). Os autores concluem que essas informações privadas podem ser usadas para aperfeiçoar serviços de plataforma personalizados e oferecer aos psicólogos sociais uma riqueza de dados que eles nunca poderiam obter de outra forma.

Identificar padrões de conduta ou de atividades inconscientemente deixadas nos (meta)dados dos sites de redes sociais cada vez mais serve para prever o futuro. Os cientistas da informação Weerkamp e De Rijke (2012) afirmam isso muito claramente: "Nós não estamos interessados nas atividades já feitas ou atuais das pessoas, mas em seus planos futuros. Nós propomos a tarefa de prever, o que envolve tentar estabelecer um conjunto de atividades que provavelmente serão populares num momento posterior". Eles situam a atividade de predição como caso especial de "mineração da vida" (*life mining*), um conceito definido como "extrair conhecimentos úteis da combinação de trilhas digitais feitas pelos indivíduos que vivem uma parte considerável de sua vida on-line". O termo "conhecimentos úteis" coloca a questão: úteis para quem? De acordo com Weerkamp e De Rijke (2012), a mídia social fornece informação significativa para que a polícia e o serviço de inteligência possam prever a atividade terrorista nascente ou planejar o controle de multidões, e para os profissionais de marketing preverem os preços futuros de ações ou potenciais receitas de bilheteria (ver também, Asur; Huberman, 2010). Das perspectivas da vigilância e do marketing, as análise preditivas – relacionando padrões de (meta)dados ao comportamento atual ou *potencial* dos indivíduos e vice-versa – produz poderosa informação sobre quem nós somos e o que nós fazemos. Quando se trata do comportamento humano, porém, essa lógica também revela uma sinuosa inclinação entre análise e projeção, e entre dedução e predição (Amoore, 2011).

Uma "mentalidade de Big Data" parece também favorecer a paradoxal premissa de que as plataformas de mídia social, ao mesmo tempo, *medem*, *manipulam* e *monetarizam* o comportamento humano on-line. Embora se acredite que os metadados retirados das plataformas de rede social refletem o comportamento humano como ele ocorre, os algoritmos empregados pelo Google, Twitter e outros sites são intrinsecamente seletivos e manipulatórios;

os usuários e os proprietários podem igualmente manipular a plataforma. Por exemplo, quando Diakopoulos e Shamma (2010) preveem preferências políticas ao analisarem o desempenho de debates por meio de *tweets*, eles parecem ignorar o potencial dos usuários do Twitter como *spin-doctors* ou militantes para influenciar os debates em tempo real no Twitter. Nos círculos de marketing, a previsão das necessidades de futuros consumidores é semelhante à manipulação do desejo: detectar padrões específicos em hábitos de consumo frequentemente resulta em tentativas simultâneas para criar demanda – uma estratégia de marketing que é monetizada com sucesso pelo famoso algoritmo de recomendação da Amazon (Andrejevic, 2011). O conteúdo da mídia social, assim como as buscas na internet, está sujeito à personalização, adaptando as mensagens a audiências ou indivíduos específicos (Pariser, 2011; Bucher, 2012). Promover a ideia dos metadados como traços do comportamento humano e das plataformas como facilitadoras neutras parece estar diretamente em conflito com as bem conhecidas *prática de filtragem* e *manipulação algorítmica de dados* por razões comerciais ou outras. A datificação e a mineração da vida se apoiam em pressupostos ideológicos, que são, por sua vez, enraizados em normas sociais dominantes. Como dito, os usuários fornecem informações pessoais às companhias e recebem serviços em troca – uma espécie de permuta. A troca de metadados por serviços de comunicação tornou-se a norma; poucas pessoas parecem dispostas a pagar por mais privacidade[3]. Seu uso como moeda para pagar pelos serviços on-line e por segurança tornou os metadados uma espécie de ativo invisível, processados, na maioria das vezes, fora de seu contexto original e sem que as pessoas tenham consciência. As companhias de mídia social monetarizam os metadados ao reprocessá-los e vendê-los para anunciantes ou companhias de dados. Os cientistas da informação, com frequência, acriticamente adotam pressupostos e pontos de vista ideológicos impulsionados pelos Serviços de Segurança Nacional e empresas de dados. O paradigma da datificação desempenha, assim, um papel profundamente ideológico na intersecção entre sociabilidade, pesquisa e comércio – um inextricável nó de funções que têm sido subavaliadas.

[3] Um relatório da Agência Europeia para a Segurança e Informação das Redes (Jentzsch; Preibusch; Harasser, 2012) mostrou que menos de um terço dos indivíduos pesquisados e um estudo sobre a troca de "dados por privacidade" estariam dispostos a pagar um valor adicional se o prestador prometesse não usar seus dados para fins de marketing.

DATAÍSMO: DESVENDANDO AS BASES IDEOLÓGICAS DA DATIFICAÇÃO

A metáfora da mineração de dados é baseada numa racionalidade peculiar que guia os empreendedores, os acadêmicos e as agências estatais na busca por um novo paradigma social-científico. Em primeiro lugar e acima de tudo, o dataísmo demonstra uma crença na objetividade da quantificação e

no potencial do rastreamento de todos os tipos de comportamento humano e da sociabilidade a partir de dados on-line. Em segundo lugar, os (meta)dados são apresentados como *matéria-prima* que pode ser analisada e processada em algoritmos preditivos sobre o comportamento humano futuro – ativos valiosos na indústria de mineração. Vamos explorar em mais detalhe cada uma dessas pressuposições ontológicas e epistemológicas que sustentam o dataísmo como um novo padrão-ouro sobre o conhecimento do comportamento humano.

Uma primeira linha de análise crítica é baseada na suposta natureza objetiva dos dados. Num provocante e reflexivo ensaio, as cientistas sociais boyd e Crawford (2012: 2) desconstroem a disseminada mitologia de que "os amplos conjuntos de dados oferecem uma forma superior de inteligência e conhecimento que pode gerar *insights* que não seriam possíveis anteriormente, com a aura de verdade, objetividade e acurácia". As pilhas de (meta)dados são propositalmente geradas a partir de diferentes plataformas on-line que são tudo menos objetivas. Os metadados relacionados ao comportamento humano agem do mesmo modo que os exames de ressonância magnética no interior do corpo: os sinais de doença nunca aparecem simplesmente na tela, mas são o resultado de cuidadosa interpretação *e intervenção* no processo de produção da imagem. Foram necessárias décadas de técnicas médicas para se aprender a representar a especificidade dos órgãos; foi necessário refinar protocolos para posicionar os corpos e ajustar o funcionamento da máquina para tornar a ferramenta mais útil (van Dijck, 2005). O Facebook e o Twitter são aparatos constantemente ajustados para traduzir amizade ou popularidade em algoritmos, promovem esses mesmos cálculos em valores *sociais* (Manovich, 2011; Bucher, 2012). Os botões de *likes* e *trending topics* podem ser muitas vezes percebidos como ícones da espontânea sociabilidade on-line, porém os algoritmos subjacentes a esses botões são sistematicamente aperfeiçoados para canalizar as respostas dos usuários (Mahrt; Scharkow, 2013).

A ideia dos (meta)dados como recursos *brutos* esperando processamento encaixa-se perfeitamente na popular metáfora da mineração de dados. De acordo com Mayer-Schoenberger e Cukier (2013), cada conjunto de dados provavelmente tem algum valor intrínseco, oculto, não revelado ainda, e as empresas estão envolvidas numa corrida para descobrir como capturar e categorizar esse valor. Mas, como de maneira acertada Gitelman (2013: 7) afirma, "dados brutos" é um oximoro: "Os dados não são fatos, eles 'são mostrados previamente ao argumento', de modo a oferecer uma base retórica. Os dados podem ser bons, maus, melhores ou piores, incompletos e insuficientes". A extração automática de dados realizada em enormes pilhas de metadados gerados pelas plataformas de redes sociais não revela mais informação sobre o comportamento humano

do que grandes quantidades de água marinha indicam sobre a poluição – a menos que você interprete esses dados utilizando métodos analíticos guiados por uma questão delimitada.

Um exemplo para ilustrar esse ponto. Uma equipe de cientistas da informação coletou durante seis meses os resultados de pesquisa, seguindo intervalos regulares de usuários que inicialmente usam a palavra-chave "financiamento residencial" em uma ferramenta de busca, de modo a descobrir correlações ao longo do tempo (Richardson, 2008). Os dados mostram como os que pesquisaram hipotecas, seis semanas após sua pesquisa inicial, vão para seguros e impostos; três meses depois eles buscam por mobiliário, e seis meses depois eles estão interessados em piscinas e acessórios para o quintal. Entretanto, correlações como esse não emergem simplesmente. Elas são muito mais induzidas por uma questão que enquadra a investigação: o que os novos proprietários de imóveis precisam comprar em seus seis primeiros meses após adquirir uma casa? Explicitar essa questão revela que um quadro interpretativo sempre prefigura a análise de dados. Seguindo a linha de pensamento de Gitelman (2013), os dados oferecem uma base retórica para o argumento de que os novos proprietários precisam de certas coisas em determinados momentos – um padrão preditivo valioso para os anunciantes.

A compreensão dos padrões requer, portanto, avaliação *crítica*: por que nós buscamos certos padrões nas pilhas de metadados, o que interessa e com quais objetivos? Identificar padrões significativos a partir dos dados retirados de plataformas on-line é um ato intrinsecamente interpretativo, ainda que as prerrogativas implícitas sejam explicitadas. As mensagens de milhões de usuárias do Facebook entre 25 e 35 anos, postando fotos de bebês em suas linhas do tempo podem ser interminavelmente analisadas em busca de padrões de comportamento, médicos ou de consumo. Os pesquisadores querem aprender sobre os hábitos alimentares de jovens mães com o objetivo de oferecer recomendações sobre mudanças? Ou pretendem descobrir padrões de necessidades de consumo para que as empresas que vendem produtos para bebês os ofereçam no momento certo? Ou, talvez com maior complexidade, são agências governamentais interessadas em interpretar esses dados para verificar sinais de depressão pós-parto ou potencial abuso infantil futuro? Os métodos quantitativos requerem firme questionamento *qualitativo* para contestar a alegação de que os padrões de dados são fenômenos *naturais*. A pesquisa de Big Data sempre envolve um (sem trocadilho) prisma explícito.

Os dados brutos não entram em uma extremidade das linhas de montagem digitais administradas pelo Google ou Facebook e a informação processada sai pela outra, como notam Mayer-Schonberger e Cukier (2013: 101). Os metadados

são pilhas carregadas de códigos multivalentes e devem ser abordados como textos sujeitos a múltiplas interpretações. De acordo com o pesquisador estadunidense John Cheney-Lippold, os dados são objetos culturais "incorporados e integrados dentro de um sistema social cuja lógica, regras e trabalho explícito de funcionamento determinam as novas condições de possibilidade de vida dos usuários" (Cheney-Lippold, 2011: 167). O Big Data configurou-se como um texto retórico que tem sido produzido para objetivos específicos e que pode ser explorado por vários grupos de pessoas, oferecendo uma alternativa para a persuasiva metáfora da mineração. Os acadêmicos, analisando os conjuntos de dados de uma perspectiva das ciências sociais ou humanas, podem colocar questões bastante diferentes das dos cientistas da informação; e médicos provavelmente veem padrões diferentes dos criminologistas (Manovich, 2011).

A lógica de convencimento do dataísmo é muitas vezes alimentada pela retórica das novas fronteiras de pesquisa, quando vastos conjuntos de dados inconscientemente ignorados, nunca acessíveis antes, abrem novas perspectivas. O dataísmo prospera partindo do pressuposto que a coleta dos dados acontece fora de qualquer estrutura predefinida – como se o Twitter facilitasse a atividade de microblog apenas para gerar dados *vivos* – e as análises de dados acontecessem sem um objetivo prévio, como se os analistas de dados analisassem aqueles dados apenas pelo interesse na acumulação de conhecimento sobre o comportamento humano. Pode não ser sempre simples identificar em qual contexto os (meta) dados são gerados e para qual propósito eles são processados. Entretanto, é fundamental explicitar as prerrogativas se os pesquisadores desejam manter a confiança dos usuários no paradigma da datificação. A confiança é parcialmente baseada na lógica persuasiva do paradigma dominante; por outro lado, a fé resiste nas instituições que transmitem a crença nos Big Data.

DATAÍSMO E CONFIANÇA NAS INSTITUIÇÕES

Uma segunda linha de escrutínio crítico encontra-se no nível das estruturas institucionais sob as quais se apoia o pensamento sobre os Big Data. As empresas de dados, agências governamentais e pesquisadores ressaltam, de maneira similar, a importância da confiança dos usuários das sociedades nas quais partes crescentes da vida cívica – desde procedimentos de petições até os registros médicos e transações financeiras – dirigem-se para plataformas on-line. Estabelecer e manter a integridade do sistema são, com frequência, tarefas atribuídas ao *estado* – no qual as plataformas têm de cumprir o conjunto de regras das agências governamentais. Quando Mayer-Schoenberger e Cukier (2013) apontam os perigos da disponibilidade de metadados ubíquos – isto é,

a representação de perfis a partir de estereótipos, as penalidades com base em propensões, a vigilância baseada na associação, um enfraquecido direito à privacidade –, responsabilizam o governo por tomar medidas que evitem esses riscos potenciais. Os autores de Big Data clamam por uma nova "classe de auditores de *big data* que chamamos algoritmistas" para "assegurar uma justa governança da informação na era dos *big data*" (Mayer-Schoenberger; Cukier, 2013: 184). Os acadêmicos também contam com os governos nacionais para regular possíveis efeitos adversos da datificação; porém eles voltam-se igualmente para as empresas de dados quando solicitam "confiança e boa vontade" das corporações e pedem que eles deem aos usuários "transparência e controle" sobre suas informações (Kosinski; Stillwell; Graepel, 2013). Na busca por confiança e credibilidade, existe uma presumível distinção entra as instituições públicas, empresariais e estatais, enquanto entidades autônomas, cada uma com uma diferente relação com os usuários – consumidores ou cidadãos.

É desnecessário dizer que nem *o estado* nem as *empresas de dados* são categorias monolíticas. Por um lado, as várias agências governamentais – além da NSA – tipificam uma relação específica com os usuários e, desse modo, assumem determinado papel na manutenção da confiança. Agências como a FTC e a NIST têm os meios legais e a obrigação política de assegurar de proteger os cidadãos contra riscos de exploração e de privacidade impulsionados pelo paradigma da datificação[4]. As empresas e companhias de dados são, por sua vez, simultaneamente concorrentes e aliadas quando o objetivo é conquistar e manter a confiança dos usuários. A crença dos usuários nas políticas de dados de uma empresa pode ser uma vantagem competitiva, mesmo assim (as parecerias nesse setor são crescentes), obrigando os usuários a manterem-se atentos sobre quem compartilha dados com quem.

No entanto, se os arquivos de Snowden nos ensinaram alguma coisa, é que é provável que as instituições que coletam e processam os Big Data não tenham se organizado *separadamente* das agências que têm a obrigação de regulá-las. Na verdade, todos os três aparatos – corporativo, acadêmico e estatal – têm apostado fortemente na obtenção de acesso irrestrito aos metadados, bem como na aceitação pública da datificação como um paradigma importante. Os cientistas, agências governamentais e as corporações, cada um por diferentes razões, têm um grande interesse nas relações mediadas por dados e no desenvolvimento de métodos que permitam a predição e a manipulação do comportamento. A aspiração de todos os agentes para saber, prever e controlar o comportamento humano se sobrepõe em certas dimensões, ainda que difira em outras. As companhias de dados desejam que suas plataformas sejam reconhecidas como objetivas, agregados padronizados de metadados – melhor e mais precisas do que as ferramentas que

[4] A agência estadunidense Federal Trade Comission (FTC) é responsável por proteger os consumidores e por impedir e prevenir práticas de negócio anticompetitivas; o National Institute of Standards and Technology (NIST) é a agência encarregada de estabelecer padrões federais de cibersegurança. Ambas as agências procuraram restaurar a confiança pública após as revelações de Snowden sobre a NSA.

as agências governamentais ou acadêmicas utilizam para medir o sentimento de consumidores, a saúde pública ou os movimentos sociais[5]. Quando as agências governamentais e os acadêmicos adotam as plataformas de mídia social como o padrão-ouro para medir o trânsito social, na verdade, eles transferem o poder sobre a coleta e interpretação de dados do setor público para o corporativo. Como boyd e Crawford (2012: 14) argumentam: "Há uma profunda diferença entre governo e indústria sobre coletar e obter o valor máximo dos dados, que leva à informação que conduz a mais publicidade direcionada, design de produto, planejamento de tráfego ou policiamento criminal".

Nesse alinhamento tripartite de forças, o governo, a academia e as empresas de dados estão interconectados no nível do pessoal, bem como pelo intercâmbio de tecnologias inovadoras, isto é, pelo codesenvolvimento de projetos de mineração de dados. Num artigo sobre o caso Snowden para o *The New York Times*, os repórteres Risen e Wingfield (2013) desnudam estreitas conexões entre o Vale do Silício e a NSA: "Ambos buscam meios de coletar, analisar e explorar grandes quantidades de dados sobre milhões de estadunidenses. A única diferença é que a NSA faz isso pela inteligência e o Vale do Silício, por dinheiro". As relações entre as empresas de dados e agências estatais de inteligência mostram como os especialistas técnicos circulam entre empregos na academia e indústrias de saúde, e transferem-se de empresas de dados para serviços financeiros ou agências de inteligência. Os interesses das corporações, da academia e das agências estatais convergem de vários modos. Por exemplo, a ferramenta Skype e sua proprietária, a Microsoft, prontamente, se envolveram com a CIA no Projeto Chess, com o objetivo de tornar as chamadas Skype utilizáveis pelos agentes da lei. Como Timothy Garton Ash (2013) ironizou num artigo de opinião no *The Guardian*: se o Big Brother viesse ao século vinte e um: "ele retornaria numa parceria público-privada".

O que está em questão aqui não é apenas a adoção do dataísmo como uma técnica de conhecimento da ação social – o comportamento humano sendo medido, analisado e previsto a partir de um amplo conjunto de metadados –, mas também enquanto uma crença na intenção das companhias de alta tecnologia e nas agências de governo de proteger os dados dos usuários de exploração. O dataísmo pressupõe *confiança* na objetividade dos métodos quantificados, bem como numa *independência* e *integridade* das instituições que utilizam esses métodos – plataformas corporativas, agências governamentais ou pesquisadores acadêmicos. A confiança e a independência são, entretanto, noções em conflito num ecossistema de conectividade no qual todas as plataformas on-line estão inevitavelmente interconectadas, tanto no nível da infraestrutura quanto no nível da lógica operacional (van Dijck, 2013b; van Dijck; Poell, 2013). Quando todas as coisas e todo mundo está conectado numa mesma infraestrutura e opera

[5] Os executivos do Google afirmam que a busca da empresa pode revelar tendências uma ou duas semanas antes que as estatísticas oficiais do governo (Bollier, 2010). Além disso, dizem que o Google Flu Trends é um instrumento melhor para medir epidemias emergentes de gripe do que os sistemas nacionais de vigilância para sintomas gripais (Wilson et al., 2009).

com a mesma lógica – uma perspectiva teorizada por Foucault bem antes do surgimento das tecnologias on-line.

Por exemplo, a lógica da análise preditiva parece ser corroborada pelos governos, pesquisadores e corporações, igualmente. O Google alega que é muito melhor que as agências estatais em *prever* estatísticas de desemprego ou de epidemias de gripe, porque seus rastreadores podem determinar quando um indivíduo está começando a procurar um novo emprego ou inicia a busca por informações sobre gripe. Os *likes* do Facebook podem *predizer* quais jovens mães podem ser suscetíveis de deixarem seus filhos desnutridos – informação sobre as quais as agências estatais de saúde podem atuar. E a NSA declara ter *prevenido* ao menos quinze ataques terroristas devido ao sistema Prism, baseado nos dados coletados de plataformas de mídia social e serviços de e-mail. O problemático nessas formas institucionais de dataísmo não está apenas no fato de faltarem esclarecimentos sobre os critérios do algoritmo utilizados para definir o que conta como uma busca por emprego, maternidade disfuncional ou terrorismo. Mais questionável é que os contextos nos quais os dados foram gerados e processador – tanto em plataformas comerciais quanto de instituições públicas – parecem ser intercambiáveis.

O que está em jogo aqui não é simplesmente nossa *confiança* em agências governamentais específicas ou em determinadas corporações, mas a credibilidade de todo o ecossistema – um ecossistema que é alimentado por um fluxo constante de bilhões de e-mails, vídeo, texto, som e metadados. A custódia sobre os fluxos de dados parece estar envolta nas dificuldades das difusas delimitações de territórios; o acesso e as restrições aos dados estão em disputa tanto no âmbito público quanto fora da zona de conhecimento das pessoas. Desde as revelações de Snowden, os cidadãos-usuários têm cada vez mais questionado as convenientes relações das empresas de alta tecnologia dos Estados Unidos com o governo e, em resposta, algumas empresas apresentaram queixas judiciais contra o que chamam de táticas de intimidação da NSA. Essa luta pública a respeito de quem os dados dos usuários devem ser confiados pode servir para aumentar a impressão de independência de cada instituição, entretanto, é óbvio que nenhuma empresa de dados, como Google e Apple, atua no vácuo. O ecossistema é tipicamente uma infraestrutura na qual nem uma única instituição está no comando (Brivot; Gendron, 2011: 153), mas cuja credibilidade é disputada em vários debates públicos, confrontos judiciais e discussões políticas – incluindo as tentativas do governo de frear os vazamentos de informantes.

A interpelação do dataísmo como crença compartilhada construída sobre a confiança institucional parece ser tão importante quanto a análise das premissas da datatificação. As noções de *confiança* e *crença* são particularmente

relevantes quanto se entende a *datavigilância* como um instrumento cada vez mais adotado para monitorar os cidadãos por meio das mídias sociais e das tecnologias de comunicação on-line (Raley, 2013). Quais são os diferentes interesses de governo, das empresas e da academia na manipulação de nossos dados? A datavigilância coloca mais questões sobre a credibilidade de todo o sistema de fluxos de informação on-line.

DATAVIGILÂNCIA E A LUTA POR CREDIBILIDADE

Alguns meses depois do desastre da NSA, o Google, o Facebook, o Yahoo e a Microsoft surpreenderam os críticos que os acusavam de ter colaborado com o governo, traindo a privacidade dos usuários, ao processar a Agência de Vigilância de Inteligência Estrangeira (Foreign Intelligence Surveillance Agency – FISA), que fornece a estrutura legal para as operações da NSA. Mark Zuckerberg do Facebook alegou, em uma entrevista de jornal, que o governo dos Estados Unidos tinha feito um "mau trabalho de equilíbrio entre a privacidade das pessoas e a tarefa de proteção", e Marissa Mayer do Yahoo admitiu que eles tinham de lutar no tribunal contra a NSA para manter a confiabilidade de sua empresa, em relação aos usuários e investidores (Rushe, 2013). Curiosamente, o que vimos como consequência das revelações de Snowden foi que as empresas de dados uniram-se e somaram forças contra a NSA para resgatar a confiança pública. A descoberta das rotineiras táticas de datavigilância ameaçou causar sério dano não apenas na confiança das pessoas nas agências estatais ou nas corporações individuais, mas no suporte ao institucional do dataísmo como um todo.

A datavigilância – o monitoramento dos cidadãos a partir de seus dados on-line – difere da vigilância em ao menos uma dimensão importante: enquanto esta presume o monitoramento para fins específicos, a datavigilância relaciona-se ao monitoramento contínuo de (meta)dados com objetivos não especificados. Portanto, a datavigilância vai bastante além do propósito de análises individuais na medida em que penetra em todo o tecido social (Andrejevic, 2012: 86). Assim, a datavigilância é uma proposta com profundas consequências para o contrato social entre as plataformas corporativas e as agências governamentais, por um lado, e os cidadãos-consumidores, por outro. Vamos olhar mais atentamente para o distinto papel de cada ator nessa batalha por credibilidade e confiança.

Desde o início, o Facebook e o Google superficialmente ancoram as expectativas de confiança de seus usuários em mantras corporativos como "Não faça mal" (Google) e "Tornando o mundo transparente e conectado" (Facebook). Para eles, o contrato social com os consumidores era baseado em tornar a sociabilidade on-line visível e rastreável; parte desse apelo por transparência

era solicitar informação pessoal e real dos seus consumidores registrados (van Dijck, 2013a). Entretanto as plataformas davam pouca transparência de volta; de 2007 até o presente momento, as empresas como o Facebook têm se envolvido em batalhas com o FTC e cortes legais para defender seus sempre mutáveis Termos de Uso, que prolongam sua política de privacidade[6]. Nos últimos anos, defensores de usuários levaram o Facebook e outras plataformas à justiça por práticas ilegais de manutenção dos *logs* de dados do usuário. Os grupos de defesa dos consumidores têm solicitado incansavelmente explicação sobre os *quid pro quo* dos serviços on-line gratuitos para ajuda a restaurar a confiança pública em plataformas específicas, bem como em todo o ecossistema. E plataformas alternativas para pesquisa e comunicação – por exemplo, Lavabit, DuckDuckGo, Path, Leaf e Silent Circle – tentam equilibrar a proteção aos dados dos usuários com serviços de qualidade. Entretanto, é muito difícil escapar das regras e práticas estabelecidas pelos jogadores dominantes no sistema.

A submissão das firmas de alta tecnologia às leis pós-Patriot Act, devidamente relatadas pelos jornalistas na sequência do caso Snowden, certamente contribuiu para a diminuição da confiança do público nas táticas de datavigilância; assim, não é surpreendente encontrar os CEOs das empresas de dados confrontando a NSA, e de viva voz tentando reestabelecer a sua imagem de facilitadores neutros. A atitude dos proprietários das plataformas em relação aos órgãos administrativos é, porém, muitas vezes ambivalente. Eles pedem aos governos para corrigir as lacunas nas leis e nas políticas (Brown; Chui; Minyika, 2011: 11), mas ao mesmo tempo as companhias alertam o governo contra a regulamentação excessiva e propõem deixar uma *abertura* a ser regulamentada pelo próprio setor de tecnologia (Schmidt; Cohen, 2010: 80).

Uma ambivalência similar vem do governo. Obviamente, as agências de inteligência têm mais interesses do que os reguladores governamentais. As questões de segurança e de privacidade com frequência colocam demandas contraditórias, lidando com definições legais ambivalentes, como a defesa de Obama dos metadados ("Nós não estamos ouvindo suas conversas telefônicas"), assim como meios legítimos para a datavigilância. Os grupos de cidadãos acertadamente reivindicam políticas claras que protejam a privacidade e a equilibrem com a segurança. Manter as definições legais em sintonia com os avançados aparatos tecnológicos é somente uma etapa fundamental no esforço de reconstruir a confiança. Como temos visto na crise bancária, iniciada em 2008, uma perda de confiança no setor financeiro foi causada por uma obscuridade semelhante, que envolveu vários esquemas financeiros complexos e a lógica de alta tecnologia dos derivativos; após duas décadas de autorregulação, a confiança no sistema bancária chegou ao nível mais baixo de todos os tempos.

[6] Durante o ano passado, o Facebook teve que defender sua prática de criar "perfis sombra" de amigos aos quais se conecta para copiar endereços e números telefônicos; a plataforma também teve que se defender de sua automática presunção de que os pais de adolescentes que usam o serviço deram a permissão para que seus nomes e imagens sejam usados em anúncios do Facebook (Oremus, 2013; Goel; Wyatt, 2013).

A responsabilidade por manter a credibilidade do ecossistema como um todo também é dos acadêmicos. O desenfreado entusiasmo de muitos pesquisadores pela datatificação como um paradigma neutro, refletindo uma crença em uma compreensão objetiva, quantificada do social, deve ser analisado de maneira mais rigorosa. A aceitação acrítica da datificação sob premissas ideológicas e comerciais pode também minar a integridade da pesquisa acadêmica em longo prazo. Para obter e conservar a confiança, os pesquisadores de Big Data precisam identificar as perspectivas parciais dos dados que analisam; mais do que manter as alegações de neutralidade, eles precisam avaliar o contexto no qual os conjuntos de dados são gerados e associar as metodologias quantitativas com indagações qualitativas. Além disso, a viabilidade e a verificabilidade da analítica preditiva como método científico merece uma investigação mais interdisciplinar, combinando, por exemplo, abordagens computacionais, etnográficas e estatísticas (Giglietto; Rossi; Bennato, 2012: 155).

Os acadêmicos são atores significativos na construção da confiança social: um paradigma incrustado nos pilares das instituições acadêmicas que, com frequência, torna-se o árbitro do que conta como fato ou opinião, como fato ou estimativa. No mundo da sociabilidade on-line, quando o comportamento humano é codificado em (meta)dados e mediado por plataformas, as distinções entre fatos, opiniões e previsões – entre objetividades, subjetividades e potencialidades – são gradualmente apagadas. Nas palavras do sociólogo Bruno Latour (2007), elas são obliteradas "de tal modo que gradativamente obtêm o mesmo tipo de visibilidade – o que não é uma pequena vantagem se nos queremos desembaraçar a mistura de fatos e opiniões que se tornou a nossa dieta informativa usual". Se a análise preditiva e as análises de dados em tempo real se tornam nossos modos preferenciais de investigação científica do comportamento humano, pesquisadores das humanidades e das ciências sociais precisam enfrentar seriamente as fundamentais questões epistemológicas e ontológicas que foram somente esboçadas nas seções anteriores.

Enquanto isso, como as ações pouco escrupulosas de Edward Snowden mostram, há um amplo ator significativo, muitas vezes esquecido, na luta pela credibilidade: os usuários-cidadãos. Quando Snowden optou por tornar pública a informação privilegiada que ele tinha sobre as práticas de datavigilância da NSA, não apenas mostrou o poder de um empregado individual para desvelar e abalar um complexo de forças estatal-industrial-acadêmico. Ele também contou com o trabalho de muitos cidadãos – pesquisadores, blogueiros influentes, jornalistas, advogados e ativistas – para tornar públicas suas preocupações a respeito de falhas estruturais no ecossistema que se desenvolve hoje. Ao longo da última década, o real poder dos usuários-cidadãos frente às plataformas

corporativas e ao estado alavancou um debate substancial, ainda que, sobretudo, em círculos de ativistas e acadêmicos. Alguns descobriram que a capacidade dos usuários resistirem às políticas de privacidade das plataformas e às estratégias de vigilância são bastante limitadas; os indivíduos encontram-se em tecnologias de plataformas e modelos de negócio de plataformas únicas cuja compreensão, em termos da interdependência e complexidade do sistema, é extremamente difícil de obter (Draper, 2012; Hartzog; Selinger, 2013; Mager, 2012). Outros pesquisadores têm defendido o letramento digital (do consumidor), particularmente num nível de entendimento da privacidade e segurança em relações aos dados sociais (Pierson, 2012). E há uma crescente massa crítica acadêmica enfatizando a importância dos usuários descobrirem como as mídias conectivas estão forjando um novo contrato social em sociedade, conforme reconfiguram a sociabilidade e a democracia em ambientes on-line (Langlois, 2013; Lovink, 2012).

O debate público muito mais amplo estimulado por Snowden é em si mesmo um forte exemplo do projeto de restaurar a credibilidade da internet[7]. É por meio de choques como esse que as pessoas tornam-se mais conscientes das forças institucionais e ideológicas envolvidas em um paradigma em evolução. A popularização da datificação como um paradigma neutro, consequência na crença no dataísmo e apoiada pelos guardiões institucionais da confiança, gradualmente produziu a perspectiva da datavigilância como uma forma "normal" de monitoramento social. Talvez isso tenha levado Snowden a denunciar essas práticas cada vez mais normalizadas, mas certamente é necessário mais do que um informante *whistleblower* para que uma análise completa sobre os novos pilares digitais da democracia e da sociabilidade seja feita. As questões colocadas na agenda por Snowden seguramente merecem permanecer nos holofotes da atenção pública até que todos os temas incertos sejam abordados. M

[7] Um levantamento de julho de 2013 do Pew Research Center for People & the Press mostrou que as revelações de Snowden têm afetado, de fato, a opinião pública a respeito da vigilância e da segurança. O relatório afirma que "a maioria dos estadunidenses (56%) diz que as cortes federais falham em oferecer limitações adequadas sobre os dados de telefone e internet coletados pelo governo como parte dos esforços antiterrorismo. Um percentual maior ainda (70%) acredita que o governo usa os dados para fins outros do que a investigação do terrorismo". Ver: <http://pewrsr.ch/1DWXOeA>.

REFERÊNCIAS

AMOORE, L. Data derivatives: on the emergence of a security risk calculus for our times. *Theory, Culture and Society*, Thousand Oaks, v. 28, n. 6, p. 24-43, dez. 2011. DOI: http://dx.doi.org/10.1177/0263276411417430

ANDREJEVIC, M. Exploitation in the Data Mine. In: FUCHS, C. et al. (Eds.). *Internet and surveillance*: the challenges of Web 2.0 and social media. Nova Iorque: Routledge, 2012. p. 71-88.

_____. The work that affective economics does. *Cultural Studies*, Abingdon, v. 25, n. 4/5, p. 604-620, 2011. DOI: http://dx.doi.org/10.1080/09502386.2011.600551

ASH, T. G. If Big Brother came back, he'd be a public-private partnership. *The Guardian*, Londres, 27 jun. 2013. Disponível em <https://goo.gl/a0Y-JZF>. Acesso em: 12 abr. 2017.

ASUR, S.; HUBERMAN, B. A. *Predicting the future with social media*. Ithaca, NY: Cornell University Library, 2011. Disponível em: <http://arxiv.org/abs/1003.5699>. Acesso em: 12 abr. 2017.

BOLLEN, J.; MAO, H.; PEPE, A. Determining the public mood state by analysis of microblogging posts. In: INTERNATIONAL CONFERENCE ON THE SYNTHESIS AND SIMULATION OF LIVING SYSTEMS, 12., 2010, Odense. *Anais…* Odense: University of Southern Denmark, 2010. Disponível em: <https://goo.gl/s3MOxM>. Acesso em: 12 abr. 2017.

BOLLIER, D. *The promise and peril of Big Data*. Washington, DC: The Aspen Institute, 2010. Disponível em: <https://goo.gl/lTEvxJ>. Acesso em: 12 abr. 2017.

BOYD, d.; CRAWFORD, K. Critical questions for Big Data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, Abingdon, v. 15, n. 5, 2012. DOI: http://dx.doi.org/10.1080/1369118X.2012.678878

BRIVOT, M.; GENDRON, Y. Beyond panopticism: on the ramifications of surveillance in a contemporary professional setting. *Accounting, Organizations and Society*, Amsterdam, v. 36, n. 3, p. 135-155, 2011. DOI: http://doi.org/10.1016/j.aos.2011.03.003

BROWN, B.; CHUI, M.; MANYIKA, J. Are you ready for the era of "big data"? *McKinsey Quarterly*, Nova Iorque, 2011. Disponível em: <https://goo.gl/l2rQs3>. Acesso em: 12 abr. 2017.

BUCHER, T. Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*, Thousand Oaks, v. 14, n. 7, p. 1164-1180, Apr. 2012. DOI: http://dx.doi.org/10.1177/1461444812440159

CHA, M. et al. Measuring user influence in Twitter: the million dollar fallacy. In: INTERNATIONAL AAAI CONFERENCE ON WEBLOGS AND SOCIAL MEDIA, 4., 2010, Palo Alto, CA. *Anais…* Palo Alto, CA: The AAAI Press, 2010. Disponível em: <https://goo.gl/ZXd308>. Acesso em: 12 abr. 2017.

CHENEY-LIPPOLD, J. A new algorithmic identity: soft biopolitics and the modulation of control. *Theory, Culture & Society*, Thousand Oaks, v. 28, n. 6, p. 164-181, 2011. DOI: http://dx.doi.org/10.1177/0263276411424420

DIAKOPOULOS, N.; SHAMMA, D. A. Characterizing debate performance via aggregated Twitter sentiment. In: CHI CONFERENCE, 28., 2010, Atlanta, GA. *Anais…* Atlanta: ACM Conference on Human Factors in Computing Systems, 2010. Disponível em: <https://goo.gl/dVtMXK>. Acesso em: 12 abr. 2017.

DING, Y. et al. Broadcast yourself: understanding YouTube uploaders. In: INTERNET MEASUREMENT CONFERENCE, 2011, Berlim. *The 2011 Internet Measurement Conference*. Disponível em: <https://goo.gl/fx7EF2>. Acesso em: 12 abr. 2017.

DRAPER, N. Group Power: discourses of consumer power and surveillance in group buying websites. *Surveillance & Society*, Abingdon, v. 9, n. 4, p. 394-407, 2012. Disponível em: <https://goo.gl/Zgb1cS>. Acesso em: 12 abr. 2017.

GIGLIETTO, F.; ROSSI L.; BENNATO, D. The Open Laboratory: limits and possibilities of using Facebook, Twitter, and YouTube as a research data source. *Journal of Technology in Human Services*, Abingdon, v. 30, n. 3-4, p. 145-159, dez. 2012. DOI: http://dx.doi.org/10.1080/15228835.2012.743797

GILLESPIE, T. The politics of "platforms". *New Media & Society*, Thousand Oaks, v. 12, n. 3, p. 347-64, 2010. DOI: http://dx.doi.org/10.1177/1461444809342738

GITELMAN, L. (Ed.). *"Raw Data" is an oxymoron (Infraestructures)*. Cambridge, MA: MIT Press, 2013.

GOEL, V.; WYATT, E. Facebook privacy change is subject of F.T.C. inquiry. *The New York Times*, Nova Iorque, 2013. Disponível em: <https://goo.gl/CiOqqv>. Acesso em: 12 abr. 2017.

HARTZOG, W.; SELINGER, E. Big Data in small hands. *Stanford Law Review Online Perspectives*, Standford, v. 66, n. 81, 3 set. 2013. Disponível em: <https://goo.gl/lUpHFh>. Acesso em: 12 abr. 2017.

HELMOND, A.; GERLITZ, C. The like economy: social buttons and the data-intensive web. *New Media & Society*, Thousand Oaks, v. 15, n. 8, p. 1348-1365, fev. 2013. DOI: http://dx.doi.org/10.1177/1461444812472322

JENTZSCH, N.; PREIBUSCH, S.; HARASSER, S.. *Study on monetizing privacy*: an economic model for pricing personal information. Heraklion: European Union Agency for Network and Information Security, 2012. Disponível em: <https://goo.gl/sdu4tG>. Acesso em: 12 abr. 2017.

LANGLOIS, G. Participatory culture and the new governance of communication: the paradox of participatory media. *Television and New Media*, Thousand Oaks, v. 14, n. 2, p. 91-105, 2013. DOI: http://dx.doi.org/10.1177/1527476411433519

LATOUR, B. Beware, your imagination leaves digital traces. *Times Higher Literary Supplement*, Londres, 6 abr. 2007. Disponível em: <https://goo.gl/doqlHK>. Acesso em: 12 abr. 2017.

LOVINK, G. *Networks without a cause*: a critique of social media. Cambridge: Polity, 2012.

KOSINSKI, M.; STILLWELL, D.; GRAEPEL, T. Private traits and attributes are predictable from digital records of human behavior. *PNAS*, Washington,

DC, v. 110, n. 15, p. 5802-5805, 2013. DOI: http://dx.doi.org/10.1073/pnas.1218772110

KWAK, H. et al. What is Twitter, a social network or a news media? In: INTERNATIONAL WORLD WIDE WEB (WWW) CONFERENCE, 19., 2010, Raleigh, NC. *Anais…* Raleigh, NC: International World Wide Web Conference Committee, 2010. p. 591-600. Disponível em: <https://goo.gl/3azQ>. Acesso em: 12 abr. 2017.

MAGER, A. Algorithmic Ideology: how capitalist society shapes search engines. *Information, Communication & Society*, Abingdon, v. 15, n. 5, p. 769-787, 2012. DOI: http://dx.doi.org/10.1080/1369118X.2012.676056

MAHRT, M.; SCHARKOW, M. The value of Big Data in digital media research. *Journal of Broadcasting & Electronic Media*, Abingdon, v. 57, n. 1, p. 20-33, 2013. DOI: http://dx.doi.org/10.1080/08838151.2012.761700

MANOVICH, L. Trending: the promises and the challenges of big social data. In: GOLD, M. K. (Ed.). *Debates in the digital humanities*. Minneapolis: University of Minnesota Press, 2011. p. 460-475.

MAYER-SCHOENBERGER, V.; CUKIER, K. *Big Data*: a revolution that will transform how we live, work, and think. Londres: John Murray, 2013.

O'CONNOR, B. et al. From Tweets to Polls: linking text sentiment to public opinion. *Association for the Advancement of Artificial Intelligence*, Palo Alto, CA, p. 122-129, 2010. Disponível em: <https://goo.gl/6FsshC>. Acesso em: 12 abr. 2017.

OREMUS, W. With friends like these: how your friends, family, and co-workers are secretly helping social networks gather intelligence on you. *Slate.com*, Brooklyn, NY, 26 jun. 2013. Disponível em: <https://goo.gl/wLbjD>. Acesso em: 12 abr. 2017.

PARISER, E. *The filter bubble*: what the internet is hiding from you. Nova Iorque: Viking, 2011.

PIERSON, J. Online privacy in social media: a conceptual exploration of empowerment and vulnerability. *Communications and Strategies*, Brussels, v. 4, n. 88, p. 99-120, 2012. Disponível em: <https://goo.gl/qcTyeq>. Acesso em: 12 abr. 2017.

QUERCIA, D. et al. Our Twitter profiles, our selves: predicting personality with Twitter. In: IEEE INTERNATIONAL CONFERENCE ON SOCIAL COMPUTING, 2011, Boston. *Anais…* Boston: IEEE ComSoc, 9-11 out. 2011. DOI: http://dx.doi.org/10.1109/PASSAT/SocialCom.2011.26

RALEY, R. Dataveillance and Countervailance. In: GITELMAN, L. (Ed.). *"Raw Data" is an oxymoron*. Cambridge, MA: MIT Press, 2013. p. 121-146.

RICHARDSON, M. Learning about the world through long-term query behavior. *ACM TWeb*, Nova Iorque, v. 2, n. 4, 2008. DOI: http://dx.doi.org/10.1145/1409220.1409224

RISEN, J.; WINGFIELD, N. Silicon Valley and spy agency bound by strengthening web. *The New York Times*, Nova Iorque, 19 jun. 2013. Disponível em: <https://goo.gl/vHizF>. Acesso em: 12 abr. 2017.

RUSHE, D.. Zuckerberg: US Government "blew it" on NSA surveillance. *The Guardian*, Londres, 11 set. 2013. Disponível em: <https://goo.gl/R5iHKk>. Acesso em: 12 abr. 2017.

SAKAKI, T.; OKAZAKI, M.; MATSUO, Y. Earthquake shakes Twitter users: real-time event detection by social sensors. In: INTERNATIONAL WORLD WIDE WEB CONFERENCE PROCEEDINGS, 2010, Raleigh, NC. *Anais…* Raleigh, NC: International World Wide Web Conference Committee, 2010. Disponível em: <https://goo.gl/QkY6na>. Acesso em: 12 abr. 2017.

SMITH, A.; BRENNER, J. Twitter Use 2012. *Pew Research Center*, Washington, DC, 31 maio 2012. Disponível em: <https://goo.gl/ZN8gcJ>. Acesso em: 12 abr. 2017.

VAN DIJCK, J. "You have one identity": performing the self on Facebook and LinkedIn. *Media, Culture & Society*, Thousand Oaks, v. 35, n. 2, p. 199-215, 2013a. DOI: http://dx.doi.org/10.1177/0163443712468605

_____. *The culture of connectivity*: a critical history of social media. Nova Iorque: Oxford University Press, 2013b.

_____. *The transparent body*: a cultural analysis of medical imaging. Seattle: University of Washington Press, 2005.

VAN DIJCK, J.; POELL, T. Understanding social media logic. *Media and Communication*, Lisboa, v. 1, n. 1, p. 2-14, 2013. DOI: http://dx.doi.org/10.17645/mac.v1i1.70

WEERKAMP, W.; DE RIJKE, M. Activity prediction: a Twitter-based exploration. *Sigir Workshop on Time-aware Information Access*. Portland, 2012. Disponível em: <https://goo.gl/6lEmGn>. Acesso em: 12 abr. 2017.

WILSON, N. et al. Interpreting Google flu trends data for pandemic H1N1 influenza: the New Zealand experience. *European Communicable Disease Bulletin*, Saint-Maurice, v. 14, n. 44, p. 429-433, 2009. Disponível em <https://goo.gl/ignVj2>. Acesso em: 12 abr. 2017.

Artigo recebido em 21 de setembro de 2016 e aprovado em 12 de janeiro de 2017.