

BIG DATA E CONTROLE DA INFORMAÇÃO NA ERA DIGITAL: TECNOGÊNESE DE UMA MEMÓRIA A SERVIÇO DO MERCADO E DO ESTADO

Ricardo M. PIMENTA

Doutor em Memória Social

Professor do Programa de Pós-Graduação em Ciência da Informação

Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT-UFRJ/ECO

Pesquisador Adjunto

Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT

ricardopimenta@ibict.br

Resumo

Este artigo é produto da pesquisa, em andamento, “dimensões tecnopolíticas do esquecimento: usos e abusos da memória na era digital” cujo objetivo é compreender as práticas e formas de memória no ambiente digital da *World Wide Web* como ação política e cultural da sociedade contemporânea. Dessa forma, neste artigo comunicar-se-ão alguns resultados preliminares no âmbito da revisão de literatura pertinente ao que é o *Big Data* e sua correlação possível com os estudos da Ciência da Informação em perspectiva interdisciplinar. Ao pensar sobre a possível “saturação” da memória através dos canais informacionais e suas respectivas tecnologias, também serão apresentados alguns dados da pesquisa de campo realizada sobre o *Big Data* na internet, enquanto lugar de produção e circulação de discursos e significações, capaz de gerar informações constituidoras de uma grande memória digitalizada da sociedade global, em muitos aspectos, criada à revelia de homens e mulheres na atualidade. Nesta primeira etapa de levantamento de dados e bibliografia, alguns resultados apontam para uma clara necessidade de se discutir o *Big Data* para além de sua dimensão econômica, refletindo sobre políticas públicas de acesso à informação e proteção à privacidade.

Palavras-chave: *Big Data*. Informação. Memória. Controle. Vigilância. Sociedade.

BIG DATA AND INFORMATION CONTROL IN THE DIGITAL AGE: TECHNOGENESIS OF A MEMORY TO SERVE MARKET AND STATE

Abstract

This article is part of the research in progress "technopolitical dimensions of forgetfulness: uses and abuses of memory in the digital age" whose goal is to understand the practices and forms of memory in the digital environment of the World Wide Web as political and cultural actions of our contemporary society. Thus, we attempt to communicate some preliminary results about the literature review concerning the subject Big Data and its possible correlations with the studies of information science in interdisciplinary perspective. Considering the possible idea of "saturation" of memory through informational channels and its respective technologies, we will also present some data from field research carried out on Big Data on Internet as a place of production and circulation of discourses and meanings that builds a large digitized memory of the global society in many ways created outside knowledge of men and women nowadays. In this first stage of data and bibliography collection and analysis, some results indicate a clear need to discuss Big Data beyond its economic dimension, reflecting on public policies for information access and privacy protection.

Key-words: Big Data. Information. Memory. Control. Surveillance. Society.

1 INTRODUÇÃO

O presente artigo é produto da pesquisa, em andamento, “Dimensões tecnopolíticas do esquecimento: usos e abusos da memória na era digital” cujo objetivo é compreender, a partir de uma perspectiva McLuhaniana, que o mesmo conjunto de tecnologias que provê a extensão de nossa memória na era digital pode também representar a própria “amputação” (MCLUHAN, 2011) de nossas capacidades de lembrar, de recuperar a informação. Dessa forma, partimos da premissa que as práticas e formas de memória no ambiente digital da *World Wide Web* constituem parte da ação política e cultural da sociedade contemporânea, marcada pelo excesso de informações através de suas respectivas tecnologias.

O termo *Big Data* surge nos últimos anos em meio a este cenário. Ele representa *grosso modo* o grande volume de dados, base para a produção de informações não estruturadas e estruturadas, produzidos de maneira exponencial na contemporaneidade. Mais do que seu volume, sua articulação em rede, sua velocidade e diversidade possibilitam a produção de mais dados, a partir dos dados já existentes, sobre indivíduos, grupos ou sobre a própria informação, quaisquer que seja ela, disponível. (BOYD; CRAWFORD, 2011).

Claramente, tal manancial de informações não somente influi direta e indiretamente nos mercados, como nos meios de vigilância já em uso pelos estados nacionais e suas instituições, produzindo registros e metainformação sobre usuários e indivíduos. Dessa maneira, indagar-se sobre as formas pelas quais atividades, informações e dados são gravados, conservados e recuperados, assim como compreender o papel das respectivas ferramentas e tecnologias empregadas é aproximar-se de elementos que interferem na memória de uma sociedade cada vez mais midiaticizada e *surveillante*.

Cabe apontar que o processo de pesquisa de campo, para o projeto citado acima, busca realizar uma espécie de etnografia do espaço da *internet*. Neste sentido, uma cartografia da *internet*, onde se busca identificar, classificar e ordenar registros presentes no ciberespaço que nos auxilie a compreender a nova relação com a memória na era digital. Dessa forma, nos alinhamos à metodologia da teoria fundamentada (GLASER; STRAUSS, 1967; GLASER, 1992), ou *grounded theory*, devido à especificidade do objeto, marcado pela atualidade, e dos fenômenos no âmbito da sociedade ainda muito recentes. Sendo assim,

este artigo trabalha com alguns dados coletados e que se situam ainda na fase inicial de “codificação aberta” proposta pela metodologia citada.

O material coletado para este artigo é parte daquele levantado na pesquisa de campo relacionada ao projeto em curso: “dimensões tecnopolíticas do esquecimento: usos e abusos da memória na era digital”.

O período exploratório da pesquisa contou com a utilização de variáveis booleanas (VANTI, 2002) na busca por identificar informações na internet por meio do uso de termos e conceitos centrais para a pesquisa como, por exemplo: “*Big Data*”, “memória”, “privacidade”, “*surveillance*”, “recuperação”, “esquecimento”, “apagamento”, “digital” e “internet”. O uso do recurso booleano visava facilitar a identificação dos registros mais pertinentes à temática analisada. Essa busca ocorreu circunscrita ao campo linguístico dos seguintes idiomas: português, inglês e francês.

Dessa maneira, ao inserir um termo como “*surveillance + information + web + privacy*” ou “*erase + information + web + forgotten*”, os endereços eletrônicos identificados eram acessados de forma a propor uma hermenêutica de seu conteúdo. Este processo está sendo realizado, como mencionado acima, também em português e francês.

Ao analisar os sítios, vídeos e demais conteúdos encontrados na *web*, passa-se para a etapa de codificação desses conteúdos. Ou seja, são identificados conceitos e categorias que classifiquem determinado fenômeno ou conteúdo significado pelo pesquisador (STRAUSS; CORBIN, 1998) com o intuito de formar categorias de análise que respondam a algumas das questões elaboradas: (1) informações referentes à competência informacional/tecnológica de apagar histórico do *browser*; (2) informações que dissessem respeito à ação de sair de redes sociais ou de serviços (apagando *login* e dados pessoais); (3) informações que se relacionassem ao debate público sobre a memória no âmbito digital, da rede mundial de computadores, que se forma e não se apaga; (4) informações pertinentes à questão legal (jurídica), ligada à privacidade e ao direito ao esquecimento; (5) informações referentes ao uso dos registros e dados que não se apagam e aos debates e aspectos do conceito de *surveillance* e tecnologia. Este artigo busca tratar deste último tópico.

A partir destas categorias torna-se possível consolidar os dados de diferentes plataformas e realizar uma análise crítica da forma como nos relacionamos na era do *Big Data* com a nossa memória na era digital e quais são os desafios em voga.

Afinal, em um cenário global margeado pelo problema do “excesso de informação”, onde o *Big Data* é reflexo desta condição, e também elemento mantenedor da própria, nos propomos a repetir a pergunta de Vaidyanathan (2011):

Estaremos nos afogando em dados, incapazes de distinguir o que é bom do que é ruim e de diferenciar o verdadeiro do falso? (...) Que ferramentas nos ajudam a lidar com essa abundância? (VAIDHYANATHAN, 2011, p. 191).

Este aspecto, mais que parte constituinte de uma memória possível da sociedade digital é, sobretudo, um estágio em plena ação da história da tecnicidade humana e, portanto, imperativo de ser compreendido para além do pragmatismo técnico-científico, de forma a exercitar uma perspectiva analítica cujo poder, cultura e formas de controle afetam desde a privacidade do homem ordinário até aquilo que se constituirá a memória de suas atividades seja no plano individual, seja no plano coletivo. Em escala global, mais dados, mais registros e mais informações acessadas pelas grandes vias de tráfego de dados tornarão experiências outrora locais, regionais e nacionais parte de uma “memória global”.

2 INFORMAÇÕES NAS MÃOS DE MUITOS ENQUANTO MAIS DADOS RESTAM NAS MÃOS DE POUCOS: A CONSTRUÇÃO DE UM PROBLEMA A SE RESOLVER

Em agosto de 2013, o Centro de Estudos sobre as Tecnologias da Informação e da Comunicação contabilizou pouco mais de três milhões e duzentos mil domínios com o registro “.br”.¹ Contudo, este número está longe de representar as dimensões da *World Wide Web*, seus territórios e vias de informação utilizadas pelos brasileiros.

De fato, há um número mais extenso de endereços eletrônicos, de dados e informações produzidas cotidianamente por todos nós. Informações em múltiplos níveis. Da *internet* à *intranet*, de compartilhamentos nas redes sociais às mensagens de texto via celulares, câmeras digitais, *tablets* e *smartphones*. A cada clique, ou *hit*, que realizamos compartilhamos informações via depoimentos, entrevistas, imagens, vídeos e documentos digitalizados. Deixamos, ademais, uma espécie de “rastro”, ou traços das atividades realizadas. Assim, temos produzido em escala exponencial cada vez mais dados por meio do uso do ciberespaço.

Dados que facilmente incorporam conjuntos de informações sobre nossas ações e representações com os quais, segundo Silva (2009), por meio do ciberespaço representam uma clara ação de extensão do “mundo da vida” (HABERMAS, 1990) que ora ajudamos a

¹ Dado extraído do sítio eletrônico <<http://cetic.br/dominios/index.htm>>. Acesso em: 7 out. 2013.

construir e a manter através do controle dos discursos e representações em jogo. Contudo, não há garantias de dispormos *grosso modo* de um fácil acesso a todos os dados que produzimos.

Mas quem, então, dispõe? É a partir deste questionamento e da tentativa de iniciarmos a construção de uma resposta, que surge a intenção deste artigo. Dessa forma, é mister compreender o capital social, político e econômico atribuído a este enorme conjunto de dados diariamente produzido no espaço eletrônico mundial, bem como levantar questionamentos sobre quais as consequências possíveis do *Big Data* para os estudos sobre economia e política da informação e de memória, respectivamente. Neste último caso, é notório que o crescente uso do *Big Data*, dissociado de políticas de preservação por meio de repositórios e arquivos capazes de fazer frente à obsolescência tecnológica, trará prejuízos tão contundentes quanto furacões e incêndios podem fazer a instalações científicas, tecnológicas e culturais como laboratórios, bibliotecas e arquivos. Tal cuidado já foi apontado por Lynch (2008) e, no recente infortúnio de um incêndio nas instalações do *Internet Archive Foundation*, na Califórnia, comprovado:

Este episódio nos fez lembrar que a digitalização e o fazer cópias são boas estratégias para o acesso e preservação. Temos cópias dos dados no *Internet Archive* em vários locais. Por isso, mesmo que o nosso edifício principal tivesse sido envolvido no fogo, nós ainda não perderíamos o conteúdo incrível que todos nós temos trabalhado tão duro para coletar (INTERNET ARCHIVE, 2013).

Para fins de preservação *lato sensu* de informações no contexto do mundo digital ou mesmo sua gestão e, portanto, o controle daquilo que é produzido e que circula no ciberespaço e em suas redes tornou-se um capital (RALEY, 2013, p. 123) a ser disputado. De economistas, médicos e biólogos a sociólogos, cientistas políticos e profissionais de *marketing*, as massivas quantidades de dados gerados em rede e acessíveis remotamente via *internet* marcam, em definitivo, um momento de grande mudança na forma como o homem contemporâneo se relacionará com a cultura, com os meios de comunicação e com os poderes que produzem conhecimento. Nesse ínterim, o fenômeno do *Big Data* não escapa à afirmação de Latour: “mude os instrumentos, e você mudará toda teoria social que os acompanha” (2009, p. 9).

São estes mesmos instrumentos que, em certa medida, operam em via de mão dupla ao passo que enquanto nos utilizamos deles, também somos monitorados e classificados por

meio dos dados por eles produzidos. O uso de contas virtuais, de transações econômicas, compras e cadastros realizados pela *internet*; todas estas atividades produzem e reproduzem dados e mais dados sobre nós. Informações capazes de, em uma perspectiva econômica (NEVES, 2006), tornar possível formas de ações objetivas das quais se configuram sistemas (HABERMAS, 1987) em prol do controle estratégico e racionalizado dos indivíduos e instituições através do avanço tecnológico capaz de recuperar e processar os dados produzidos diariamente.

É notório que o termo *Big Data*, iniciado como forma de identificar um conjunto complexo de dados em volume e velocidade extraordinários, acabou sendo apropriado, e legitimado, pelo mercado global e hoje, com o avanço do campo da Tecnologia da Informação (TI), se faz presente de forma cada vez mais horizontalizada em um vasto campo de atividades não somente econômicas, mas científicas e políticas. Dessa forma, afirmamos que o controle da informação advinda do uso do *Big Data* é exemplo claro não somente do controle estratégico e racionalizado das sociedades, como posto logo acima, como da reificação da “virada informacional” proposta por Lopes (2008, p. 23), onde, do ponto de vista econômico, “a produção de riqueza passa a depender do conhecimento e informações produzidos” (LOPES, 2008, p. 71).

Não obstante, torna-se evidente o potencial político do controle do *Big Data* quando compreendemos que seu advento representa um momento, ou uma experiência de “concretização” de uma mudança tecnológica. Neste sentido, a ação do processamento do *Big Data* por governos e empresas nos remete, senão ao próprio conceito de “individualização técnica”, proposto por Simondon (1969), por meio da qual se estende, pela tecnologia empregada, essa concretização da capacidade do objeto técnico: dados. Capazes de gerar metadados que, conectados à rede mundial de computadores, podem ser processados de maneiras diversas de forma a produzirem mais e mais informações.

Em concordância com proposição de Barthes (1970, p. 140-141), afirmamos que as novas “catedrais” da era digital são projetadas não apenas pela “neomania”, também presente na “febre” do *Big Data*, mas pelo maciço investimento nas Tecnologias de Informação e Comunicação (TICs) e em seu consumo. São as novas catedrais feitas de *bytes*, algoritmos, programações. Mas quais as possíveis implicações desse processo?

3 VIGILÂNCIA E INFORMAÇÃO COMO FORMAS DE CONSTRUÇÃO DE METAMEMÓRIAS

É sabido que o espaço público tornou-se o cenário de profusão de dados em escala jamais vista e, da mesma forma, que é justamente a partir deles que novos dados são produzidos: os metadados. Hoje, cada vez mais, homens e mulheres são comumente traduzidos em um conjunto de dados em meio ao extenso universo da linguagem — *numérique* — binária. Tudo isso é passível de ser encontrado em estudos e análises de *Big Data* realizados com objetivos econômicos ou políticos. Afinal, empresas, organizações de tipos diversos e governos são cada vez mais capazes de se informar das ações referentes à localização, deslocamento e consumo dos indivíduos com precisão mais e mais sofisticada. Contudo, ainda são carentes as pesquisas sobre o que o *Big Data* acarreta e quais as implicações sociológicas, históricas e políticas a respeito de sua exploração.

Certamente, produzimos diariamente dados sobre nós mesmos e sobre nossas atividades e relações comerciais, políticas e até mesmo privadas. Produzimos algumas vezes sem saber, como um ingênuo jovem que navega na *internet*, um histórico de nosso trânsito, de nosso consumo, de nossos gostos e declarações. E isso acarreta na possível construção de um novo suporte da memória por meio das atuais tecnologias de vigilância e monitoramento.

Este novo “suporte” é capaz de informar com exatidão algumas de nossas ações mais ordinárias, como o tempo e o exato lugar que ocupamos no deslocamento de nossas casas ao trabalho através do sistema de transporte público. Em Paris, por exemplo, o uso do cartão *Navigo* pelos usuários da rede de transportes RATP (*Régie Autonome des Transports Parisiens*) possibilita claramente tal monitoramento e, portanto, o controle de uma “memória artificial” de nosso deslocamento no espaço público da cidade.

De fato, manter o anonimato tornou-se, cada vez mais, uma prerrogativa difícil a ser mantida por aqueles que utilizam o *Navigo*; carregado pelo uso do cartão bancário e que se torna necessário a cada entrada ou troca de transporte, seja ele metrô, trem, ônibus, bicicletas ou *tramways*. A frase “*circulez, vous êtes fichés!*” (MANACH, 2009), algo como “circule, você está fichado!”, publicada no *Le Monde* em 2009, aponta claramente para a ameaça ao direito ao anonimato e o cenário de vigilância que extrapola o direito de ir e vir em prol do monitoramento arbitrário. Uma nova forma de vigilância e, portanto, de controle está lançada. Implementada, como política pública de transporte, mas também de segurança pública. E porque não, de informação?

De Paris a Londres, de Nova Iorque ao Rio de Janeiro, mesmo as atividades da vida cotidiana têm se tornado grandes fontes de informação para o mercado e para o Estado. Tornamo-nos “sensores humanos” graças ao uso de aparelhos inteligentes que, conectados à “nuvem”, nos tornam emissores de dados em tempo real com uma velocidade inequívoca. Este fenômeno muda, portanto, nossa paisagem tecnológica e influencia nossas ações em dimensão social, cultural, econômica e política.

A grande massa de dados que produzimos, sem necessariamente estarmos completamente conscientes disso, é um dos elementos fulcrais para as novas formas de vigilância e de controle. (ROUVROY, 2010). Realizados na interseção do público e do privado, esses dados são a evidência de uma “infraestrutura de informação” (SÁ, 2011, p.7) que, ao fazer uso da perspectiva de Candau (1996; 1998), afirmamos ser elementos constituidores de uma metamemória.

Esta infraestrutura possível não se faz com os dados somente, mas por meio de sistemas capazes de conjugá-los com o objetivo de capitalizá-los em torno de projetos de mercado, segurança pública, controle de tráfego, rastreamento, entre outros. Obviamente, em muitos casos, os dados podem ser empregados em campos diversos, como os mencionados acima, sob a linguagem estatística. Não por acaso, este “instrumento por excelência do Estado” (BOURDIEU, 2012, p. 337) capaz de concentrar informações, processá-las e produzir conhecimento exclusivo e inacessível ao indivíduo, ao cidadão, pode servir como ferramenta catalisadora do volume de dados produzidos.

Nesse sentido, o controle de dados não apenas facilita a “normalização da vigilância” (DANDEKER, 1990), como torna visível o cidadão ordinário que, através de sua interação com as novas TICs, acaba por prover dados passíveis de serem recolhidos, armazenados e analisados em escala muito além da privacidade esperada por ele.

Esses dados, quando organizados e utilizados por quaisquer instituições públicas ou privadas, acabam por se tornar um capital capaz de servir a fins diversos dentro de uma economia informacional onde são eles, como já mencionamos, os dados, o capital em questão cujo controle e organização alimentam uma “memória artificial” aplicada.

E-mails, mensagens de texto, tuítes ou simples navegação na *web* deixam traços digitais (MENCHEN-TREVINO, 2013, p. 2). Registros numéricos e códigos concentrados em *gigabytes*, *terabytes* e até *zettabytes* detêm potenciais informações de todos nós capazes de interferir no cotidiano de nossas vidas de maneira inequívoca uma vez que, entre o real e o

virtual, encontramos-nos cada vez mais entre mundos tecnologicamente mediados (BALL; WEBSTER, 2003).

A sociedade da informação na era digital é, portanto, uma sociedade marcada pela produção, circulação, armazenamento e controle de uma quantidade massiva de dados. Dados que no presente, apesar de sua variedade de plataformas e linguagens, respectivamente de produção e tradução, e apesar do monopólio da língua inglesa no tocante à “infraestrutura global de informação” (FERREIRO, 1997, p. 238), convergem para uma única forma de linguagem, digitalizada, binária, da qual não somos capazes de extrair quaisquer informações sem as ferramentas computacionais apropriadas e suas competências necessárias.

A diferença está aí. Se ambas, a linguagem natural e a documentária, são sistemas simbólicos constituídos com o fim de auxiliar no processo comunicacional (CINTRA et al., 2002), apesar de suas singularidades presentes, o *Big Data* possui possivelmente uma característica dupla: é resultado da explosão informacional vivida na atualidade, mas é também solução para o desafio ainda atuante da recuperação da informação.

Este desafio, contudo, não é exclusividade do *Big Data*. Saracevic (1996) já havia identificado que, no tocante à recuperação da informação, esta figurava como o principal problema a ser explorado pela Ciência da Informação, uma vez que as formas como se realiza a recuperação estão implicitamente ligadas às tecnologias da informação (TIs) com as quais lidamos; e, através das quais mudaram dramaticamente as maneiras como procedemos (SARACEVIC, 2007, p. 1916).

Ainda assim, com o desenvolvimento dos usos do ciberespaço por indivíduos, instituições e estados, juntamente com suas tecnologias respectivas, a explosão informacional parece ter se mostrado detentora de uma capacidade exponencial. E, portanto, o desafio apenas cresce. Na era do *Big Data*, o manancial de dados detém um potencial de produção de informações com o qual não sabemos se estaremos aptos a conjugá-los harmoniosamente.

Atualmente, o processamento de todos esses dados tem colocado a questão da informação e sua gestão como alguns dos bens mais essenciais à comunicação, à identificação e ao controle de pessoas e suas respectivas atividades no espaço público, apesar de não se restringir apenas a ele. E isto pode ser um problema.

A biometria, por exemplo, associada aos cartões de identificação (*ID-cards*) compostos por *chips* nos quais estão contidas múltiplas informações sobre seu portador, poderá ser capaz de substituir o *labor* da memória individual; além de outros tantos mecanismos de recuperação de informações que operam por mediações e que interferem na forma como transitamos pelo espaço público e acessamos seus diferentes lugares, seja de trabalho, de lazer ou mesmo de passagem. O Brasil, por exemplo, já possui um projeto de Registro de Identidade Civil (RIC), nos moldes dos *ID-cards*, que se encontra em fase de implantação e testes.

Figura 1 - Imagem de vídeo de explicação das tecnologias e aplicações da RIC



Fonte: PORTAL BRASIL, 2013.²

Através desta nova forma de mediação entre indivíduo e sociedade, o *ID-card* juntamente com o próprio corpo se tornarão os repositórios da informação. E o indivíduo ficará desobrigado de lembrar-se de senhas e nomes, códigos ou doses. Basta ser ele mesmo e o aparelho de leitura biométrica saberá por ele ao passo que conseguirá lê-lo. Mas quais os possíveis problemas inerentes a essa tecnologia, capaz de prover tamanha acessibilidade? Como na imagem acima, a proposta de um “documento com segurança” não deve ser

² Disponível no sítio eletrônico “Portal Brasil” em <<http://www.brasil.gov.br/cidadania-e-justica/2010/12/conheca-o-novo-registro-de-identidade-civil-ric>>. Acesso em: 20 nov. 2013. A imagem foi recolhida no perfil público do autor, utilizando o DIIGO: ferramenta de *Personal Knowledge Management* (PKM). Disponível em: <<https://www.diigo.com/item/image/1a1qj/mdxu>>. Realizada em: 20 nov. 2013.

encarada com truísmo. A segurança, como uma forma de controle, opera em uma via de mão dupla. É parte da inteligência informacional que historicamente se recrudesciu em meio às relações político-institucionais características da sociedade (THIESEN, 2011, p. 8).

Liberdades civis, como o direito à privacidade, e direitos humanos (KHAN, 2006) poderão estar em xeque com os usos escolhidos para tais tecnologias. E afinal, apesar de Lovink (2011, p. 302) considerar as dicotomias público/privado e global/local obsoletas graças às novas formas de apropriação e suas estratégias no ambiente de rede, as ditas “clássicas” dicotomias não nos parecem estar completamente solucionadas. Mais ainda, em um cenário de disputa como o da *internet*, os atores sociais e seus antagonistas encaram o mesmo desafio de não se perderem no “mar” de dados gerados pelos diversos espaços de comunicação e suas linguagens no ciberespaço.

Parece certo afirmar que a memória que se construirá no futuro dependerá em escala crescente dos dados preservados, disponibilizados e controlados, pois é fato que nem tudo se disponibiliza. Dados estes que, por sua vez, serão dependentes da tecnologia utilizada. Dessa forma, apesar da crença de que os suportes tecnológicos poderão *grosso modo* nos salvar do blecaute informacional, eles também submeterão nossa capacidade de lembrar, uma vez que o ato de rememoração estará cada vez mais imbricado à própria tecnologia, à luta contra a obsolescência tecnológica. E dessa maneira, conclui-se que o receio de que toda a tecnologia de análise de informações acabe por propiciar o advento de uma “ditadura dos dados” (CUKIER; MAYER-SCHÖNBERGER, 2013), é pertinente, pois o que se registra, organiza e usa dependerá, em estágios diferentes, ainda que subordinados uns aos outros, da tecnologia vigente cada vez mais “autônoma”.

4 DOMAR OS DADOS: O QUE SELECIONAR E O QUE APAGAR NA CULTURA DO “SALVAMENTO”

Segundo McAfee e Brynjolfsson (2012), a cada segundo mais dados transitam pela *internet*, do que nos últimos 20 anos. Se tal afirmação for minimamente próxima da realidade, enfrentamos um desafio jamais visto no tocante à gestão da informação produzida hoje e celebrada pela cultura informacional e pela tecnicidade de nossas relações com o mundo, sendo no espaço público ou privado.

Os dados em sua profusão representam um novo domínio da sociedade contemporânea na qual suas instituições de controle têm como desafio a operacionalização e, portanto, a “racionalização” (HABERMAS, 1973, p. 27) com o intuito de atingir

determinados fins. Os fins, no entanto, são menos comuns em seus objetivos enquanto se assemelham assustadoramente em seu *modus operandi*. Econômicos ou políticos, estes fins acabam passando, em muitos aspectos, pelo controle da informação. Capaz de gerar dividendos, além de ações e reações, em ambos os cenários público e privado da sociedade civil, a informação potencializada pelo advento da *World Wide Web* tornou-se inegavelmente um capital e sua economia um “mundo sem regras” (JOHN, 2000, p. 130).

Mas o que deve ser guardado, o que pode ser compartilhado e a quem é permitido o acesso são questões que definem claramente os limites políticos do controle da informação advindo de um manancial de dados tão extenso e igualmente profundo como o *Big Data*.

Volume, velocidade e variedade (três Vs) são características do *Big Data*. O controle do que se usa em termos de dados e informação requer, portanto, capacidade de arbitrar de forma objetiva sob o risco de perder-se no intenso fluxo de dados gerados a cada momento. Vejamos que: “o pesadelo da informação insuficiente que fez nossos pais sofrerem foi substituído pelo pesadelo ainda mais terrível da enxurrada de informações que ameaça nos afogar” (BAUMAN, 2011, p. 8).

O *Big Data* representa nada mais que esta “enxurrada” contra a qual parece se precisar aprender a sobreviver. E no campo da ciência, por exemplo, ela está progressivamente mais e mais presente nas pesquisas em curso. No caso do laboratório do Centro Europeu de Pesquisas Nucleares (CERN), responsável pelo maior acelerador de partículas do mundo, controlar os três “Vs” requer selecionar o que descartar.

[...] muitas vezes, somos obrigados a escolher dados para guardar e outros para descartar. Para armazenar aqueles 15 *petabytes* por ano, o CERN escolhe dados relevantes entre 15 *petabytes* gerados por segundo de operação do LHC. Como saber o que guardar e guardar cada vez mais rápido é o desafio (XEXÉO, 2013, p. 19).

De fato, selecionar é, sobretudo, uma forma de controle na qual a gestão da informação e do conhecimento é necessária. Segundo Souza, Dias e Nasiff, ela faz parte do

entendimento de como as pessoas, a informação e o conhecimento se relacionam dinamicamente, em detrimento de programas e/ou modelos gestores fundamentados em noções e soluções tecnológicas, num movimento em direção aos processos de conhecer (SOUZA; DIAS; NASIFF, 2011, p. 56).

Compreender seus dados e saber como usá-los e controlá-los parece ser o maior desafio na atualidade quando o assunto é *Big Data*. Entretanto, para esses desafios serem vencidos é necessário, antes de tudo, obter acesso a este conjunto tão extenso, profundo e

heterogêneo de dados. Ter acesso é ter permissão. É possuir a chave de entrada para uma espécie de “cofre” onde os *bytes* são o capital. Sendo assim, é óbvio que os dados, provenientes de nossas ações e produções cotidianas, influenciarão a imagem, os gostos e preferências, os pontos positivos e negativos, nosso perfil e, por que não, nossa identidade no mundo à frente. Isso é memória. E ela deixou de ser nossa propriedade quando passamos a torná-la produto de consumo (HUYSEN, 2000) e partilhá-la via mundo eletrônico.

Destacamos aqui a recente conferência realizada por Chomsky no evento “*Engaging Data 2013*”, realizado no dia 13 de novembro de 2013. Para ele, “qualquer sistema de poder, seja ele representado pelo Estado ou por grandes empresas como o *Google*, se utiliza dos recursos tecnológicos possíveis para controlar e maximizar seu poder” (CHOMSKY, 2013 *apud* HATCHIMONJI, 2013).

De fato, o emprego na tecnologia necessária para tal é evidente. É possível averiguar, por exemplo, que a busca pelo termo “*Big Data*” obteve um crescimento exponencial na *internet* durante os últimos dois anos.

O gráfico apresenta um número relativo ao total de pesquisas (eixo y), referentes àquele termo, realizadas no espaço de tempo (eixo x). Dessa forma, não são números absolutos. Ainda assim, é possível averiguar a tendência quanto ao interesse por determinado termo na *internet*³.

Figura 2 - Imagem de gráfico que mostra a incidência do termo *Big Data* no cenário mundial, de 2004 a 2013

³ Mais informações sobre o *Google Trend* em <<https://support.google.com/trends/answer/87285?hl=pt-BR>>. Acesso em: 28 nov. 2013.



Fonte: Extraído do Google Trends. Acesso em: 11 nov. 2013.⁴

O aumento do interesse no processamento de colossais conjuntos de dados não é por acaso, isso é certo. Sua entrada é claramente visível pelo termo “*Big Data*”. Como o próprio serviço do *Google* sugere, é possível identificar uma curva ascendente, com alguns picos, em torno do termo buscado no *Google*, nos últimos anos. Há de fato um aumento da busca pelo termo, sugerindo um crescente conjunto de informações sobre *Big Data*. Esta curva, no entanto, mostra apenas o que está sob o campo de cobertura do motor de busca *Google*.

O *Big Data* na era digital é, talvez, a constatação de nossa pulsão pelo controle da informação cibernética (WIENER, 1984), uma vez que a linha é cada vez mais tênue quando pensamos nos limites da influência e controle que exercemos e a qual estamos submetidos em relação aos dados que produzimos e suas respectivas tecnologias capazes de interpretá-los. De fato, o volume de informações que indivíduos são capazes de produzir, sobre eles próprios em redes sociais como *Facebook*, *Twitter*, *Instagram*, *Orkut*, *LinkedIn*, entre outras, certamente já compõe o maior banco de dados sobre o cotidiano das pessoas; no qual emoções, preferências profissionais, de consumo, políticas ou sexuais são registradas em um grande histórico que integra uma memória global (ASSMANN; CONRAD, 2010), comunicativa (ASSMANN, 2011, p. 17-19), plural e perigosamente efêmera no tocante aos discursos e experiências compartilhadas de uma geração.

⁴ A imagem foi recolhida no perfil público do autor, utilizando a ferramenta de PKM, DIIGO. Disponível em: <<https://www.diigo.com/item/image/1a1qj/syw5>>. Elaborada em: 11 nov. 2013.

Esta memória, portanto, não nos pertence completamente. E aí reside um possível problema. Das redes sociais às páginas comerciais dispostas na *internet*, nos parece correta a interpretação do conceito de “informação assimétrica” (AKERLOF, 1970), segundo Vila, Greenstadt e Molnar (2003, p. 3) e quanto à forma como nossos dados serão tratados. Para eles, se o usuário comum da *web* não possui ferramentas necessárias para avaliar qual tratamento receberão seus dados, as companhias responsáveis por guardá-los não se sentirão cobradas ou pagas suficientemente para zelar por eles. E isso significa não zelar pela privacidade. Nesse sentido, poder-se-ia esperar que a privacidade não seja respeitada em quaisquer sítios eletrônicos.

Aliás, ela não deve ser de todo respeitada uma vez que os insumos necessários para o uso do *Big Data* são em diversos casos, dados que nenhum de nós sabe plenamente que disponibilizamos. E isto por si só já aponta para um difícil e improvável respeito à privacidade quando o campo é a *internet*.

5 CONCLUSÃO

Se há, portanto, um obstáculo à privacidade, há um igual desafio ao segredo. Mas, o que é segredo na *internet*, que seu provedor de *e-mail* não saiba? Que a empresa responsável pela rede social, na qual optamos por criar um perfil, não esteja ciente? Mesmo quando deletamos um perfil nessas redes, saibamos, não há garantias de que nossos dados foram também apagados em seu *data center*. E o mesmo ocorre com o provedor de *e-mail*. Dados jamais apagáveis? Gleick (2013, p. 405) afirma que “a informação produzida e consumida pela humanidade costumava desaparecer”. De certa maneira, esperava-se que desaparecesse. Algo tão incontestável quanto a tempestade que horrorizava o *Angelus Novus* de Benjamin (1985), anjo que representava a história, parece hoje ser posto em xeque pela tecnologia criada pelo homem. Tecnologia esta que ainda produz mais dados em estado caótico que informações organizadas e objetivadas. Mas que anda a passos largos para um controle dos mesmos.

Este controle será a nova “penicilina” contra a distopia informacional que “adoece” a sociedade do conhecimento na era digital, e da qual todos nós detemos uma parcela de responsabilidade. Mas como qualquer grande vacina, tratamento, técnica, ou saber, há a concentração desse novo capital “tecno-informacional”, seguida de sua exploração. Pensar sobre quais serão seus desdobramentos é, portanto, fundamental.

REFERÊNCIAS

- AKERLOF, George A. The market for "Lemons": quality uncertainty and the market mechanism. **The Quarterly Journal of Economics**, v. 84, n. 3, p. 488-500, Aug. 1970.
Disponível em:
<<https://www.iei.liu.se/nek/730g83/artiklar/1.328833/AkerlofMarketforLemons.pdf>>.
Acesso em: 25 nov. 2013.
- ASSMANN, Aleida; CONRAD, Sebastian (Ed.) **Memory in a global age. Discourses, practices and trajectories**. New York: Palgrave Macmillan Memory Studies, 2010.
- ASSMANN, Jan. Communicative and cultural memory. In: MEUSBURGER, Peter; HEFFERNAN, M.; WUNDER, E. (Ed.) **The geographical point of view**. Dordrecht: Springer, 2011. Series: Knowledge and Space, v. 4.
- BALL, Kirstie; WEBSTER, Frank (Ed.) **The intensification of surveillance: crime, terrorism and warfare in the information era**. London, UK: Pluto Press, 2003.
- BANDEIRA-DE-MELLO, R.; CUNHA, C. Grounded theory. In: GODOI, C. K.; BANDEIRA-DE-MELLO, R., SILVA, A. B. d. (Ed.). **Pesquisa qualitativa em estudos organizacionais: paradigmas, estratégias e métodos**. São Paulo: Saraiva, 2006.
- BARTHES, Roland. **Mythologies**. Paris: Seuil, 1970.
- BAUMAN, Zygmunt. **44 cartas do mundo líquido moderno**. Rio de Janeiro: Jorge Zahar Editor, 2011.
- BENJAMIN, Walter. Sobre o conceito de História. In: _____. **Obras escolhidas. Magia e técnica, arte e política**. São Paulo: Brasiliense, 1985.
- BOURDIEU, Pierre. **Sur l'état. Cours au Collège de France (1989-1992)**. Paris: Éditions Raisons d'agir; Seuil, 2012.
- BOYD, Danah; CRAWFORD, Kate. Six provocations for big data. In: A DECADE IN INTERNET TIME: SYMPOSIUM ON THE DYNAMICS OF THE INTERNET AND SOCIETY. **Proceedings...** Oxford: Oxford Internet Institute's, 21 set. 2011. Disponível em:
<http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431>. Acesso em: 26 mar. 2014.
- CANDAU, Joël. **Anthropologie de la mémoire**. Paris: PUF, 1996.
- _____. **Mémoire et identité**. Paris: PUF, 1998.
- CINTRA, Ana Maria et al. **Para entender as linguagens documentárias**. 2. ed. São Paulo: Polis, 2002.
- CUKIER, Kenneth; MAYER-SCHÖNBERGER, Victor. **Big Data: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana**. Rio de Janeiro: Editora Campus, 2013.
- DANDEKER, Christopher. **Surveillance, power and modernity: bureaucracy and discipline from 1700 to the present day**. New York: St. Martin's Press, 1990.
- FERREIRO, Soledad. Dealing with diversity and digital culture. **International Information & Library Review**, v.29, n.2, p.237-245, June 1997.
- GLASER, Barney. **Basics of grounded theory analysis**. Mill Valley, CA: The Sociology Press, 1992.

GLASER, Barney; STRAUSS, Anselm. **The discovery of grounded theory: strategies for qualitative research**. New York: Aldine Transaction, 1967.

GLEICK, James. **A informação: uma história, uma teoria, uma enxurrada**. São Paulo: Cia das Letras, 2013.

HABERMAS, Jürgen. **The theory of communicative action**. Boston: Beacon Press, 1987. v. 2: Lifeworld and system: a critique of functionalist reason.

_____. **Pensamento pós-metafísico: estudos filosóficos**. Rio de Janeiro: Tempo Brasileiro, 1990.

_____. **La technique et la science comme "ideologie"**. Paris: Éditions Gallimard, 1973.

HATCHIMONJI, Grant. Chomsky, Gellman talk Big Data at MIT Conference. **CSO Online**, 15 Nov. 2013. Disponível em: <<http://www.csoonline.com/article/743285/chomsky-gellman-talk-big-data-at-mit-conference>>. Acesso em: 23 nov. 2013.

HUYSEN, Andreas. **Seduzidos pela memória: arquitetura, monumentos, mídia**. 2. ed. Rio de Janeiro: Aeroplano, 2000.

INTERNET ARCHIVE. Fire Update: Lost Many Cameras, 20 Boxes. No One Hurt. **Internet Archive Blog**, 6 Nov. 2013. Disponível em: <<https://blog.archive.org/2013/11/06/scanning-center-fire-please-help-rebuild/>>. Acesso em: 21 nov. 2013.

JOHN, Nancy R. The ethics of the click: users and digital information in the internet age. **Libri: International Journal of Libraries and Information Services**, v. 50, p. 129–135, 2000. Disponível em: <<http://www.librijournal.org/pdf/2000-2pp129-135.pdf>>. Acesso em: 19 nov. 2013.

KHAN, Arfan. Identity cards: the final nail in the coffin of civil liberties? **Journal of Criminal Law**, n. 70, p. 139-146, 2006.

LATOURE, Bruno. 'Tarde's idea of quantification'. In: CANDEA, M. **The social after Gabriel Tarde: debates and assessments**. London: Routledge, 2009. Disponível em: <<http://www.bruno-latour.fr/sites/default/files/116-CANDEA-TARDE-FR.pdf>>. Acesso em: 23 mar. 2014.

LOPES, Ruy Sardinha. **Informação, conhecimento e valor**. São Paulo: Radical Livros, 2008.

LOVINK, Geert. Atualizando a mídia tática. Estratégias de midiativismo. In: ALBAGLI, Sarita; MACIEL, Maria Lucia (Org.). **Informação, conhecimento e poder: mudança tecnológica e inovação social**. Rio de Janeiro: Garamond, 2011.

LYNCH, Clifford. How do you data grow? **Nature**, v.455, n.7209, 4 Sept. 2008. Disponível em: <<http://www.nature.com/nature/journal/v455/n7209/full/455028a.html>>. Acesso em: 23 nov. 2013.

MCAFEE, A; BRYNJOLFSSON, E. Big Data: the management revolution. **Harvard Business Review**, Oct. 2012. Disponível em: <<http://hbr.org/2012/10/big-data-the-management-revolution>>. Acesso em: 20 nov. 2013.

MCLUHAN, Marshall. **Understanding media: the extensions of man**. Critical Edition by W. Terrence Gordon. 2. ed. Berkeley, California: Gingko Press, 2011.

MANACH, Jean-Marc. **Le passe Navigo "anonyme" n'existe pas**. Paris [s.n.] 2009. Disponível em: <<http://bugbrother.blog.lemonde.fr/2009/01/07/le-passe-navigo->

[%E2%80%9Canonyme%E2%80%9D-nexiste-pas/](#)>. Acesso em: 14 out. 2013. O blog “Bug Brother: qui surveillera les surveillants?” faz parte da seção de *blogs* convidados do portal eletrônico do jornal Le Monde.

MENCHEN-TREVINO, Ericka. Collecting vertical trace data: big possibilities and big challenges for multi-method research. **Policy & Internet**, v.5, n.3, 2013. Disponível em: <<http://www.ericka.cc/vertical.pdf>>. Acesso em: 2 nov. 2013.

NEVES, Marcelo. **Entre Têmis e Leviatã**: uma relação difícil: o Estado democrático de direito a partir e além de Luhmann e Habermas. São Paulo: Martins Fontes, 2006.

PORTAL BRASIL. Conheça o novo Registro de Identidade Civil (RIC) [cidadania e justiça], 18 ago. 2013. Disponível em: <<http://www.brasil.gov.br/cidadania-e-justica/2010/12/conheca-o-novo-registro-de-identidade-civil-ric>>. Acesso em: 20 nov. 2013.

RALEY, Rita. Dataveillance and counterveillance. In: GITELMAN, Lisa (Ed.) «**Raw Data**» is an **oxymoron**. Cambridge, Massachusetts; London, England: MIT Press, 2013.

ROUVROY, Antoinette. La “digitalisation de la vie même”: enjeux épistémologiques et politiques de la mémoire digitale. **Documentaliste – Sciences de l’Information**, v. 47, n. 1, p.63-64, 2010. Disponível em: <http://www.adbs.fr/adminsite/objetspartages/liste_fichiergw.jsp?OBJET=DOCUMENT&CODE=76971425&LANGUE=0>. Acesso em: 13 out. 2013.

SÁ, Alberto Manuel Teixeira de. **Arquivos dos media e preservação da memória: processos e estratégias do caso português na era digital**. Braga: Universidade do Minho, 2011. Tese (Doutorado) – Ciências da Comunicação, Instituto de Ciências Sociais, Universidade do Minho, Braga, 2011. Disponível em: <<http://repositorium.sdum.uminho.pt/bitstream/1822/19638/1/Alberto%20Manuel%20Teixeira%20de%20S%C3%A11.pdf>>. Acesso em: 9 out. 2013.

SARACEVIC, Tefko. Relevance: a review of the literature and a framework for thinking on the notion in Information Science. Part II: Nature and manifestations of relevance. **Journal of the American Society for Information Science and Technology**, v. 58, n.13, p. 1915-1933, Nov. 2007.

_____. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>>. Acesso em: 9 nov. 2013.

SILVA, Nilza Pereira da. O ciberespaço como extensão do mundo da vida. **Revista Espaço Acadêmico**, n. 97, jun. 2009. Disponível em: <<http://periodicos.uem.br/ojs/index.php/EspacoAcademico/article/viewFile/7116/4142>>. Acesso em: 14 out. 2013.

SIMONDON, Gilbert. **Du mode d’existence des objets techniques**. Paris: Aubier-Montaigne, 1969.

SOUZA, Edivanio Duarte de; DIAS, Eduardo José W.; NASSIF, Mônica Erichsen. Informação e do conhecimento na Ciência da Informação: perspectivas teóricas e práticas organizacionais. **Informação & Sociedade**: estudos, João Pessoa, v.21, n.1, p. 55-70, jan./abr. 2011. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/4039/5598>>. Acesso em: 20 nov. 2013.

Tendências da Pesquisa Brasileira em Ciência da Informação, v. 6, n. 2, jul./dez. 2013.

STRAUSS, Anselm L.; CORBIN, Juliet M. **Basics of qualitative research: techniques and procedures for developing grounded theory**. 2. ed. London: SAGE Publications, 1998.

THIESEN, Icléia. “Inteligência informacional” e Ciência da Informação: um esboço de trajeto. **LIINC em Revista**, v. 7, n. 1, p. 6-18, mar. 2011. Disponível em: <<http://revista.ibict.br/liinc/index.php/liinc/article/viewFile/400/260>>. Acesso em: 26 mar. 2014.

VAIDHYANATHAN, Siva. **A googlelização de tudo (e por que devemos nos preocupar)**: a ameaça do controle total da informação por meio da maior e mais bem-sucedida empresa do mundo virtual. São Paulo: Cultrix, 2011.

VANTI, Nadia Aurora Peres. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação na difusão do conhecimento. **Ciência da Informação**, v. 31, n. 2, p.152-162, 2002. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/171/150>>. Acesso em: 26 mar. 2014.

VILA, Tony; GREENSTADT, Rachel; MOLNAR, David. Why we can't be bothered to read privacy policies: models of privacy economics as a Lemon Market. In: **INTERNATIONAL CONFERENCE ON ELECTRONIC COMMERCE, 5. Proceedings...**[S.l.] ACM, 2003. p.403-407. Disponível em: <<http://www.dmolnar.com/papers/econprivacy.pdf>>. Acesso em: 25 nov. 2013.

WIENER, Norbert. **Cibernética e sociedade**: o uso humano de seres humanos. São Paulo: Cultrix, 1984.

XEXÉO, Geraldo. Big Data: computação para uma sociedade conectada e digitalizada. **Ciência Hoje**, v. 306, p.19-23, ago. 2013. Disponível em: <http://cienciahoje.uol.com.br/revista-ch/2013/306/pdf_aberto/bigdata306.pdf>. Acesso em: 17 nov. 2013.