RIO DE JANEIRO 2016

Big Data no Sul Global

Relatório sobre estudos de caso

ITS
Instituto
de Tecnologia
& Sociedade
do Rio

Introdução

É comum, na literatura sobre o assunto, identificar fases do desenvolvimento de regulamentações de proteção de dados. De fato, Viktor Mayer- Schöemberger costumava identificar gerações de leis de proteção de dados, a primeira delas visando a regulamentar os poucos grandes computadores capazes de gerenciar bancos de dados no fim dos anos 1960 e início dos anos 1970, e a quarta (a última documentada por este autor) visando a construir um ambiente eficaz para a aplicação das leis, com autodeterminação informativa individual e a assistência de autoridades em proteção de dados. A regulamentação da proteção de dados, no entanto, enfrenta uma crise em alguns de seus elementos mais centrais e tradicionais – e essa crise provavelmente levará tal regulamentação, e a própria governança de dados pessoais, a uma nova geração.

Conceitos como o livre consentimento individual e o princípio de finalidade, que um dia foram o eixo central das leis de proteção de dados, estão pouco a pouco passando para uma posição complementar à medida que cada vez mais dados pessoais não mais são coletados diretamente dos indivíduos, ou são sem o seu conhecimento. O Big Data, junto com outras inovações como as relacionadas com a Internet das Coisas (Internet of Things – IoT), estão mudando sobremaneira o cenário referente aos dados pessoais e exigirão que as agências regulatórias se adaptem a essas novas circunstâncias, seja adequando as ferramentas e os princípios existentes, seja criando novos meios para a aplicação das leis.

Preocupações sobre Big Data e proteção de dados, obviamente, também estão presentes no Sul Global. No caso do Sul Global, ainda que estejamos lidando com

tecnologias de alcance mundial que, com muita frequência, são fornecidas por agentes globais com características similares no mundo inteiro, também devem ser consideradas algumas questões particulares. Os países no Sul Global são basicamente consumidores, e não fornecedores, das tecnologias que estruturam o Big Data, o que, em teoria, pode fazer com que essas tecnologias não sejam tão adequadas às suas necessidades específicas.

Além disso, o Big Data e sua cadeia de consequências – classificação social, processos decisórios baseados em algoritmos, e assim por diante – podem ser usados para produzir discriminação ou mesmo tornar ainda mais fortes as barreiras econômicas e sociais na sociedade em alguns países. E isso tem mais probabilidade de acontecer em países que não possuem uma regulamentação adequada sobre assuntos relacionados à proteção de dados e direitos digitais, como é o caso de muitos países da região, ou quando o cenário político pode se beneficiar de tecnologias para construir sistemas de controle e vigilância por razões políticas. Portanto, ainda que estejamos lidando com um paradigma tecnológico (Big Data) que consideraríamos global, há razões para supor que sua implementação no Sul Global deve ser analisada segundo alguns critérios específicos.

Também é relevante o fato de que as revelações de Edward Snowden posicionaram o Brasil, a maior economia da América do Sul, no meio do escândalo Snowden, depois que alguns documentos revelaram que a Agência de Segurança Nacional dos Estados Unidos (NSA) estava grampeando os telefonemas da presidenta brasileira.[1] Não foi por acaso que o Brasil, juntamente com a Alemanha – outro país cuja chefe de Governo foi vítima de escuta telefônica –, apresentou uma proposta de resolução sobre privacidade online à Assembleia Geral da ONU, que foi aprovada por unanimidade[2] e levou à produção de um relatório, por parte do Alto Comissariado de Direitos Humanos da ONU, sobre o direito à privacidade na era digital e à criação de um mandato para um Relator Especial sobre o Direito à Privacidade,[3] nomeado recentemente.[4]

Embora o uso de Big Data no Brasil seja relativamente recente, o volume de dados processados no Brasil (e no Hemisfério Sul) é tão impressionante quanto o do Hemisfério Norte. No entanto, o Brasil – diferentemente de outros países do Sul, mesmo latino-americanos – não possui uma lei geral de proteção de dados, como veremos neste relatório. Este volume imenso de dados sendo processados também é consequência do uso, por usuários locais, de ferramentas e plataformas web disponíveis globalmente, como Google, YouTube e Facebook. Apenas para dar um

[1] Ver http://www.globalpost.com/dispatch/news/regions/americas/brazil/130709/us-spying-brazil-snowden-leaks

[2] UN GA Resolution 68/167. The right to privacy in the digital age. Available at http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/68/167. Accessed 25 November 2015.

[3] Ver http://www.ohchr.org/EN/Issues/Privacy/SR/Pages/SRPrivacyIndex.aspx. Accessed 25 November 2015.

[4] Ver http://www.ohchr.org/EN/Issues/Privacy/SR/Pages/JoeCannataci.aspx. Accessed 15 December 2015.

exemplo, o Facebook armazena, em média, 111 MB de informações de usuários[5] e, de acordo com um ex-presidente do Google, Eric Schmidt, 5 exabytes são criados diariamente na internet, uma quantidade que corresponde a todo o volume de informações geradas por nossa civilização desde o seu surgimento até 2003.[6]

Além da coleta de dados por meio do uso de ferramentas e plataformas web, dados também estão sendo coletados e armazenados por meio de uma grande variedade de ferramentas tecnológicas. Para dar um exemplo, apenas na cidade do Rio de Janeiro "estima-se que existam cerca de 700 mil câmeras instaladas em ruas, edifícios, condomínios, bancos, supermercados etc., que, de algum modo, registram nossa vida cotidiana."[7] Como alguns autores destacam, "estamos deixando cada vez mais vestígios de nossa vida diária."[8]

Nesse cenário, uma série de iniciativas usando Big Data também floresceu no hemisfério sul. Um grande número de conjuntos de dados estão sendo publicados online por governos que seguem os princípios de governo aberto; startups e empresas bem estabelecidas estão à procura de novos dados para promover novos aplicativos, ou para manipular informações a fim de patrocinar produtos lucrativos; até mesmo os cidadãos estão gerando Big Data ao usar cada vez mais aplicativos, websites e software para manter seus registros na nuvem. Certamente, o Big Data representa várias oportunidades, mas essa tendência não está livre de preocupações. Entre as áreas de preocupação com relação ao Big Data e a possíveis danos e oportunidades, podemos citar aplicação das leis, segurança, saúde pública, transporte, direitos do consumidor e outras. O Big Data promete benefícios a uma sociedade mais aberta, mas também apresenta riscos, alguns já familiares (como a proteção da privacidade) e outros, novos. Um motivo de preocupação no uso de Big Data é o fato de que "o acesso a tantos dados, de tantas fontes diferentes, e à capacidade computacional necessária para processá-los, cada vez mais significa que podemos perceber padrões, participar de descobertas e revelar segredos que até então estavam ocultos". [9]

Considerando este cenário, o presente relatório apresentará dois estudos de caso realizados pelo Instituto de Tecnologia e Sociedade sobre o uso de Big Data para fins

---

[5] Ver https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/privacidade_em_tempos_de_big_data?lang=en

[6] Ver http://techcrunch.com/2010/08/04/schmidt-data/. Accessed 25 November 2015.

[7] Ver https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/privacidade_em_tempos_de_big_data?lang=en

[8] Ver, por exemplo, LEMOS, André. Cibercultura e Mobilidade: a Era da Conexão. Razón y Palabra, N. 41, Oct/Nov 2004. Available at http://www.razonypalabra.org.mx/anteriores/n41/alemos.html

[9] KUNER, Christopher et at. The challenge of 'big data' for data protection. Editorial. International Data Privacy Law, 2012, Vol. 2, No. 2.

de fiscalização no Brasil, um sobre a Polícia Federal e o outro sobre a Receita Federal e o Banco Central. A análise começará definindo o escopo do que deve ser considerado Big Data, e então focará nas preocupações advindas do uso de Big Data.

Em seguida, apresentaremos a análise dos dois estudos de caso, destacando as melhores práticas, os benefícios e os possíveis impactos de cada estudo de caso. A análise também considerou o possível impacto da adoção de uma lei geral de proteção de dados no Brasil.

Sumário

1/
O que é Big Data?

Um dos jargões mais ubíquos dos últimos anos é "Big Data". Este conceito, um produto quase natural do desenvolvimento cada vez mais rápido da tecnologia e dos modelos de negócio dela dependentes, tem se mostrado de valor inestimável aos mais variados campos, da pesquisa acadêmica à análise de negócios, e até mesmo em políticas públicas. Mas o que é Big Data?

Simplificando, podemos dizer que Big Data é, literalmente, o conjunto de dados cuja existência só é possível em consequência da coleta massiva de dados que se tornou possível nos últimos anos, graças à onipresença de aparelhos e sensores na vida cotidiana e do número crescente de pessoas conectadas a tais tecnologias por meio de redes digitais e também de sensores. Todas as ações e comunicações em plataformas digitais, como com telefones celulares, computadores ou mesmo transações de cartão de crédito e, mais recentemente, declarações de imposto de renda, ou ações que, em algum momento, são digitalizadas e assim transformadas em dados, como as câmeras de segurança associadas com software de reconhecimento facial ou de padrões[10], são passíveis de serem armazenadas, processadas, copiadas e distribuídas quase instantaneamente, possibilitando análises de dados que podem levar governos e empresas a tomar decisões supostamente melhor fundamentadas.

Embora parte desses dados seja coletada sem o consentimento expresso dos sujeitos sobre quem essas informações são geradas (como no caso de registros de passageiros em voos internacionais), grande parte deles é disponibilizada e fornecida pelos próprios sujeitos – por meio do uso de redes sociais, compras online e basicamente tudo que fazem online que possa estar conectado com sua identidade.[11] Entretanto, tais procedimentos de prospecção e análise de dados em grande escala, apesar dos possíveis benefícios que podem trazer, causam muitas preocupações, e aquelas relacionadas com privacidade e proteção de dados pessoais estão no topo da lista. Como se tem mostrado de tempos em tempos, na era do "Big Data", qualquer coisa que você diz e faz pode, e provavelmente será usado, contra ou a seu favor.[12]

---

[10] Ver "UK, the world's most surveilled state, begins using automated face recognition to catch criminals" http://www.extremetech.com/extreme/186435-uk-the-worlds-most-surveilled-state-begins-using-automated-face-recognition-to-catch-criminals

[11] Ver o plano da China para criar um sistema obrigatório de classificação de cidadãos baseado em redes sociais e no comportamento de consumo https://www.aclu.org/blog/free-future/chinas-nightmarish-citizen-scores-are-warning-americans.

[12] Um exemplo desse possível uso de informações disponíveis publicamente é a intenção do Departamento de Segurança Interna dos Estados Unidos de incorporar publicações em redes sociais em suas análises para concessão de vistos. Ver http://www.theverge.com/2015/12/14/10124498/homeland-security-social-media-visa-review.

2/ Preocupações

Tem-se afirmado que quanto mais informações sobre indivíduos são coletadas por diferentes agentes em diferentes contextos, mais se enfraquece a autonomia individual: muitas vezes não consentimos expressamente que nossos dados sejam coletados (por exemplo, no contexto de análise de dados de navegação na internet) e, mesmo quando nos é dada a opção de usar um serviço e ter nossos dados coletados ou desistir de usar o serviço em questão, essa escolha é feita por meio do uso de políticas de privacidade e termos e condições praticamente ilegíveis.[13] De acordo com Tene e Polonetsky (2012)[14]:

Claramente, a coleta de grandes conjuntos de dados e o uso de analítica implica questões de privacidade. As tarefas de garantir a segurança dos dados e de proteger a privacidade se tornam mais difíceis à medida que as informações se multiplicam e são compartilhadas pelo mundo de forma ainda mais disseminada. Informações sobre a saúde dos indivíduos, sua localização, seu consumo de eletricidade e suas atividades online são minuciosamente examinadas, causando preocupações relacionadas com classificação, discriminação, exclusão e perda de controle.

Essas informações não só estão sendo coletadas com pouca ou nenhuma noção de consentimento, como frequentemente são tratadas sem qualquer responsabilidade ou transparência pública, e compartilhadas, vendidas e transmitidas a terceiros. Mesmo no caso de conjuntos de dados anonimizados propositadamente, mostrou-se[15] que a re-identificação pode anular muitas tentativas de de-identificação, insinuando que até mesmo Big Data "anônimo" ainda possui muitas questões relacionadas com privacidade. As preocupações com privacidade são especialmente alarmantes quando se lida com coleta de dados e vigilância em massa por parte do governo: como mostram os incontáveis escândalos envolvendo a Agência de Segurança Nacional dos Estados Unidos (NSA) e o Quartel-General de Comunicações do Governo britânico (GCHQ), quando não fiscalizado, o poder do Estado pode ser tão ou mais nocivo para a liberdade e a privacidade do indivíduo que as iniciativas privadas. Embora o uso de Big Data pelo Estado possa, sem dúvida, levar a mais eficiência e mesmo a mais segurança, é vital que tanto as etapas relacionadas com os dados propriamente ditos – coleta, manuseio, análise etc. – quanto as decisões tomadas com base nesses dados sejam submetidas a transparência e escrutínio público,

[13] Ver o artigo de Alexis Madrigal para o The Atlantic: "Reading the Privacy Policies You Encounter in a Year Would Take 76 Work Days" http://www.theatlantic.com/technology/archive/2012/03/reading-the-privacy-policies-you-encounter-in-a-year-would-take-76-work-days/253851/

[14] Ver Tene, O.; Polonetsky, J., "Privacy in the age of Big Data: a time for big decisions" http://www.stanfordlawreview.org/sites/default/files/online/topics/64-SLRO-63_1.pdf

[15] Ver Felten, E. W.; Huey, J.; Narayanan, A. "A Precautionary Approach to Big Data Privacy" http://randomwalker.info/publications/precautionary.pdf. Acessado em 15 de dezembro de 2015.

e sejam executadas com o máximo cuidado para com a privacidade daqueles com cujos dados o governo está lidando. Entretanto, isso quase nunca acontece, como ilustraremos com os dois estudos de caso sobre iniciativas realizadas pelo governo brasileiro. Em ambas as análises apresentaremos a implementação do sistema, como suas práticas estão relacionadas com o conceito de Big Data e, finalmente, seus possíveis benefícios e implicações sobre a privacidade e a proteção dos dados.

Focando nesses dois tipos de preocupação, escolhemos como estudos de caso duas situações em que o governo brasileiro faz uso de Big Data. Ambas são soluções que não estão bem documentadas e que tendem a desconsiderar preocupações acerca de transparência e proteção de dados, apesar dos interesses da administração pública e do argumento tecnocrático de eficiência. Em sua descrição, alguns de seus possíveis malefícios serão revelados.

3/

Sistemas de Registro de Identificação de Passageiros (PNR) e de Informações Antecipadas sobre Passageiros (API)

O primeiro desses sistemas consiste, na verdade, de dois sistemas relacionados: o Registro de Identificação de Passageiros (PNR) e as Informações Antecipadas sobre Passageiros (API). Implementados pela Agência Nacional de Aviação Civil (Anac) em 2012 por meio de sua Resolução 255,[16] esses sistemas funcionam em paralelo, obrigando as companhias aéreas[17] a armazenarem e transmitirem uma ampla gama de informações sobre cada voo internacional, passageiro e tripulante entrando e saindo ou simplesmente fazendo escala ou conexão em território brasileiro. As informações coletadas devem ser transmitidas eletronicamente para o Departamento de Polícia Federal (DPF) antes de cada voo – portanto, os dados pessoais dos passageiros são coletados antes de o voo chegar ao Brasil. Oficialmente, o sistema visa a "prevenção e a repressão a atos de interferência ilícita", incluindo, por exemplo, evasão fiscal sobre

[16] Ver Resolução ANAC 255/12: http://www2.anac.gov.br/transparencia/audiencia/aud22_2012/3%20-%20Resolucao%20-%20API%20e%20PNR%20(versao%20final).pdf

[17] Com a exceção de helicópteros e jatos particulares ou usados em serviços de táxi aéreo.

produtos importados ou criminosos procurados, bem como a facilitação dos processos migratórios em vários níveis burocráticos. Também há planos de ampliar o marco de segurança ao qual esses sistemas pertencem, por exemplo com a adoção de tecnologia de reconhecimento facial biométrico em aeroportos, capaz de comparar os rostos dos passageiros com um banco de dados de indivíduos de "alta periculosidade".[18]

Além disso, sistemas PNR também foram implementados ou propostos nos Estados Unidos[19] e na União Europeia[20] e em mais de cinquenta outros países, e são recomendados pela Associação Internacional de Transporte Aéreo (IATA).[21]

[18] Se Ver "Receita Federal lança declaração eletrônica de bens de viajantes": http://www.receita.fazenda.gov.br/AutomaticoSRFsinot/2013/08/16/2013_08_16_13_09_50_734484890.html

[19] Ver http://www.cbp.gov/document/forms/passenger-name-record-pnr-privacy-policy

[20] Ver https://euobserver.com/justice/130430

[21] Ver "ANAC determina regras sobre repasse de dados de passageiros à Polícia Federal" http://economia.uol.com.br/ultimas-noticias/infomoney/2012/11/19/anac-determina-regras-sobre-repasse-de-dados-de-passageiros-a-policia-federal.jhtm

3.1/ Práticas de Big Data nos sistemas PNR e API

Somente em 2013, 19,2 milhões de passageiros passaram pelo Brasil em voos internacionais, um número que, ao que tudo indica, continua crescendo.[22] Os sistemas PNR e API se propõem a coletar e armazenar informações privadas, sensíveis e detalhadas de cada um desses passageiros (bem como o número de malas e o peso da bagagem).

Especificamente, mais de quinze campos de informação por passageiro, incluindo nome completo, gênero, endereço domiciliar e de cobrança, data e local de nascimento e até mesmo informações de cartão de crédito. Isso, por si só, constitui uma grande quantidade de informações, e o fato de que tais informações são compartilhadas ativamente com outras agências, possibilitando, assim, a referência cruzada com outros bancos de dados específicos, é um exemplo típico de uso de Big Data por parte do setor público.

[22] Ver "ANAC divulga Anuário do Transporte Aéreo de 2013": http://www.anac.gov.br/Noticia.aspx?ttCD_CHAVE=1584

3.2/ Benefícios

Os sistemas PNR e API visam a facilitar o trabalho das agências do governo. Ao coletar informações pessoalmente identificáveis sobre cada passageiro passando pelo Brasil e unificar o acesso a bancos de dados hoje dispersos, as agências do governo são capazes de pesquisar passageiros automaticamente em listas de suspeitos ou criminosos procurados, ajudar a detectar evasões fiscais e acelerar o processo de entrada no país – isso é especialmente verdadeiro tendo em vista a implementação de um sistema para a declaração eletrônica de bens do viajante (a chamada e-DBV).[23]

E, sem dúvida, quando não se consideram as implicações para a privacidade ou a proteção de dados, é fácil entender de que modo mais dados sobre os passageiros podem ajudar as investigações, sejam criminais ou não.

[23] Ver "Receita Federal lança declaração eletrônica de bens de viajantes"
http://www.receita.fazenda.gov.br/AutomaticoSRFsinot/2013/08/16/2013_08_16_13_09_50_734484890.html

3.3/ Possíveis implicações

Praticamente todos os tipos de informação que são classificados como Informação Pessoalmente Identificável sob a definição usada pelo Instituto Nacional de Padrões e Tecnologia do Departamento de Comércio dos Estados Unidos estão presentes no PNR. Portanto, tanto o volume quanto a natureza dos dados coletados são sensíveis o suficiente para requerer um sistema de proteção mais robusto. Entretanto, embora na licitação[24] que a Polícia Federal brasileira abriu para contratar empresas para a implementação do PNR conste resumidamente que tais sistemas devem estar em conformidade com suas diretrizes de segurança[25], estas são, quando muito, genéricas: não se encontrou nenhuma informação a respeito de pontos cruciais, e além disso surgem várias outras preocupações com questões sobre privacidade após a análise do escasso referencial jurídico que as ampara. Alguns pontos importantes são apresentados a seguir, e muitas vezes coincidem com aqueles listados pela IATA como importantes ao implementar sistemas PNR[26]:

Limitação de finalidade: a natureza dos dados coletados é bastante pessoal. Do gênero do passageiro à sua data de nascimento e ao seu número de cartão de crédito, coletam-se informações privadas e sensíveis, sem que a sua necessidade seja clara. Embora a segurança pública seja de importância fundamental, o acúmulo compulsório de dados pessoais de milhões de indivíduos sem maior controle ou critério é feito às expensas da privacidade pessoal, quando não atende a um critério de proporcionalidade. Até mesmo a eficácia dessa medida é questionável quando se considera seu caráter seletivo: conforme se mostrou, a transmissão de tais dados só é compulsória para o público em geral – e não para aeronaves particulares, o que induz igualmente à presença de um elemento discriminatório.

Autorização para acesso aos dados: embora conste que os dados de PNR e API devem ser transmitidos e armazenados "com segurança", não fica claro quem, isto é, quais agências e quais funcionários, efetivamente terão autorização para acessar esses dados. Não está claro se as conexões com outros bancos de dados serão feitas automaticamente por algum tipo de algoritmo, ou manualmente, e nem mesmo quais bancos de dados estão de fato conectados com tais sistemas. Há observações de que a Secretaria da Receita Federal (SRF) brasileira tem acesso aos registros do PNR, e também de que estes podem ser usados para facilitar a obtenção de autorização para acesso aos dados junto a autoridades de saúde pública e de controle de substâncias (como a Anvisa), mas

[24] Ver http://www.dpf.gov.br/servicos/licitacoes/2013/distrito-federal/orgaos-centrais/cgti/pregoes/pregao-eletronico-no-07-2013-cgti-dpf

[25] Ver Federal Police Department's "Portaria. No.779/2009-DG/DPF"

[26] Ver IATA "Guidelines on Passenger Name Record (PNR) Data" https://www.iata.org/iata/passenger-data-toolkit/assets/doc_library/04-pnr/New%20Doc%209944%201st%20Edition%20PNR.pdf

não se sabe exatamente como tal acesso é feito e controlado.

Duração do período de armazenamento das informações: a velocidade de crescimento desses dados é enorme. Seu armazenamento por um período de tempo indefinido não parece se justificar, mesmo se considerarmos os supostos objetivos de vigilância e detecção de evasão de impostos. No entanto, em momento algum há referência ao período de tempo durante o qual os dados de PNR e API teriam de ser mantidos.

Divulgação de dados individuais: os indivíduos cujas informações são coletadas devem ser capazes de obter uma cópia dessas informações, tanto por questões de transparência quanto para que possam fazer alterações em tais dados. Embora isso constitua, em geral, um direito constitucional segundo a legislação brasileira[27] e inclusive possa ser obtido por meio de uma solicitação de acesso à informação[28], em nenhum ponto da normativa esse direito é descrito.

Sem acesso aos próprios dados o indivíduo não tem possibilidade de retificar possíveis inexatidões nos dados coletados que podem ter consequências negativas para si mesmo.

[27] Os cidadãos têm o direito de entrar com pedido de habeas data nos tribunais brasileiros, um tipo de procedimento por meio do qual podem requerer acesso a informações sobre si mesmos mantidas pelo governo.

[28] A legislação brasileira sobre Acesso à Informação (Lei 12.527 de 2011) pode ser usada por indivíduos para solicitar suas informações pessoais, de acordo com seu procedimento para solicitação de informações.

4/

Sistemas da Receita Federal e do Banco Central: CCS, SCR e referência cruzada

O segundo conjunto de sistemas analisados é composto de vários bancos de dados e configurações de referência cruzada usados pela Secretaria da Receita Federal (SRF) e pelo Banco Central do Brasil (BCB) para fins de controle fiscal. Por meio de um marco regulatório complexo, as autoridades fiscais têm uma gama enorme de fontes nas quais se basear durante suas investigações, que podem ser divididas em duas categorias:

A. Rastreamento do sistema financeiro: isso inclui vários bancos de dados mantidos pelo Banco Central e pela SRF que coletam informações sobre operações bancárias. Entre eles, estão:

I. Cadastro de Clientes do Sistema Financeiro Nacional (CCS): introduzido em 2003 pela Lei 10.701, que estabeleceu mudanças na legislação preexistente sobre lavagem de dinheiro, o CCS permite ao Banco Central manter registros de cada conta bancária no país, contendo dados sobre i. identificação dos clientes, ii. instituições onde eles mantêm contas ou ativos e iii. datas de início e término de cada relação comercial. Isso não inclui informações bancárias sensíveis detalhando transações ou movimentações de conta.

II. Sistema de Informações de Crédito do Banco Central (SRC): um banco de dados centralizado detalhando cada operação de crédito que exceda mil reais. O sistema armazena clientes inadimplentes ou não, e atribui a cada um deles uma classificação de crédito com base em seu histórico de bons ou maus pagadores.

III. Declarações ou notificações de terceiros também são um método comum: por exemplo, por meio da Instrução Normativa RFB Nº 811,[29] a SRF determina que os bancos devam enviar

[29] Ver http://normas.receita.fazenda.gov.br/sijut2consulta/link.action?visao=anotado&idAto=15765

relatórios mensais detalhados sobre transações financeiras de todas as contas que excedam 5 mil reais (para pessoa física) ou 10 mil reais (para pessoa jurídica) em um período de seis meses. Essa declaração é chamada Declaração de Informações sobre Movimentação Financeira (Dimof), e se baseia normativamente na Lei Complementar n. 105/2001, regulamentada pelo Decreto Presidencial 4489/2002.[30] Há outras declarações similares, emitidas por empresas de cartão de crédito (Decred), imobiliárias (Dimob) e até mesmo do setor de saúde (Dmed)[31].

B. Referência cruzada com bancos de dados externos: o governo também usa integração com outros bancos de dados, bem como monitoramento de redes sociais, durante suas atividades de auditoria fiscal:

1. Análise de redes sociais: o secretário da Receita Federal, Jorge Rachid, declarou no ano passado[32] que o órgão monitora rotineiramente os perfis dos contribuintes nas redes sociais em busca de incongruências entre suas declarações de renda e suas informações disponíveis publicamente. Até o momento, não se sabe se tal prática é institucionalizada ou se é feita esporadicamente, e não se encontrou nenhum texto normativo sobre a mesma.

II. Acesso a vários bancos de dados de registro de propriedade: a SRF também tem acesso direto ou indireto a vários outros bancos de dados e fontes de informação. Em particular, obtém acesso ao Registro Nacional de Veículos Automotores (Renavam),[33] ao Registro Aeronáutico Brasileiro (RAB)[34] e ao Registro de Propriedade Marítima (PRPM) (em conjunto com autoridades portuárias).[35] Internacionalmente, também há um acordo de cooperação para troca de informações com autoridades norte-americanas.[36]

---

[30] Ver http://www.planalto.gov.br/ccivil_03/decreto/2002/D4489.htm

[31] Ver http://exame.abril.com.br/seu-dinheiro/noticias/os-dedos-duros-que-entregam-quem-burla-o-imposto-de-renda

[32] See http://www.correiobraziliense.com.br/app/noticia/ir2015/2015/05/01/interna2-ir0215,481607/receita-federal-monitora-redes-sociais-dos-contribuintes.shtml

[33] Ver http://normas.receita.fazenda.gov.br/sijut2consulta/link.action?visao=anotado&idAto=20559

[34] Ver http://www.anac.gov.br/Noticia.aspx?ttCD_CHAVE=137

[35] Ver http://exame.abril.com.br/seu-dinheiro/noticias/os-dedos-duros-que-entregam-quem-burla-o-imposto-de-renda

[36] Ver http://idg.receita.fazenda.gov.br/noticias/ascom/2015/agosto/acordo-brasil-eua-permitira-troca-de-informacoes-sobre-contribuinte

4.1/ Práticas de Big Data nos sistemas da Receita Federal e do Banco Central

Embora a quantidade e a variedade de fontes utilizadas pela SRF certamente aproxime suas técnicas dos métodos comuns de Big Data, não está claro se todos esses sistemas e bancos de dados estão interconectados por computadores ou se são verificados manualmente e separadamente, de acordo com demandas atípicas.

Entretanto, o fato de que a SRF intensificou seus esforços tanto para acumular (por um aumento no número de declarações fiscais diferentes) quanto para acessar dados, aliado a seus grandes investimentos em tecnologia,[37] sinaliza, no mínimo, uma tendência à adoção de práticas de Big Data na auditoria fiscal.

[37] Recentemente, por exemplo, 15 milhões de reais foram gastos em software de reconhecimento facial em aeroportos.

Ver http://agenciabrasil.ebc.com.br/economia/noticia/2014-09/receita-reforcara-fiscalizacao-de-passageiros-de-voos-internacionais-em

4.2/ Benefícios

Sem dúvida, o acesso a mais dados de modo geral, e a dados mais confiáveis em particular, contribui para que o trabalho de auditoria fiscal e de combate ao crime da Secretaria da Receita Federal e do Banco Central do Brasil seja muito mais completo e eficiente. A capacidade de detectar automaticamente (isto é, computacionalmente) pessoas físicas e jurídicas que cometem evasão fiscal, em vez de confiar no trabalho manual de comparação e análise de papelada, torna a auditoria fiscal mais rápida e ágil.

4.3/ Possíveis implicações

A eficiência, entretanto, não é um fim em si mesma, e não pode dirimir a importância dos direitos à privacidade e ao sigilo bancário, sobretudo quando, como neste caso, tais práticas são feitas com pouca ou nenhuma transparência ou responsabilidade. O recurso ao Big Data deve estar acompanhado de considerações sobre suas implicações jurídicas e também éticas, ou corre o risco de afetar de tal maneira o equilíbrio de poder entre o cidadão e a administração pública que pode impactar inclusive a natureza democrática de uma sociedade. Portanto, alguns aspectos sobre as possíveis implicações do uso de Big Data nos sistemas da Receita Federal e do Banco Central são apresentados a seguir:

A. Transparência: Embora os muitos tipos de declaração fiscal existentes sejam bem regulamentados – por natureza, já que impõem claras obrigações de informação a terceiros –, muitas das outras práticas que envolvem acesso a bancos de dados são mal documentadas, quando o são, e descritas apenas de forma marginal em veículos de comunicação ou comunicados de imprensa. Com a exceção do acesso da SRF a registros de propriedade de veículos, não se encontra nenhuma menção ao acesso oficial a bancos de dados de aeronaves e embarcações no corpus normativo[38] da SRF. Não se conhece, por exemplo, o volume total de dados pessoais sendo retransmitidos desses bancos de dados para a SRF. Mais preocupante, no entanto, é a declaração de que as redes sociais são vigiadas por auditores fiscais. Não há absolutamente nenhuma informação a esse respeito nas fontes oficiais, o que indica que tal prática é feita de maneira não oficial, assistemática, e contra os princípios de legalidade e publicidade (ver artigo 37 da Constituição da República brasileira).

B. Limites do recurso aos dados: Não se conhece o alcance total da vigilância que está sendo realizada – por exemplo, se esta só faz uso de informações disponíveis publicamente, se tais informações são armazenadas e posteriormente usadas como evidência ou simplesmente como um indício que leva a uma investigação oficial documentada etc. Ainda a respeito de declarações fiscais, também há debate sobre a constitucionalidade de várias medidas adotadas pela SRF em particular e pelo governo federal em geral, especificamente com relação à Lei Complementar n. 105/2001. Essa Lei concede ao Executivo poder discricionário para acessar registros bancários e financeiros sem supervisão judicial e, ainda que, em teoria, isso possa ser no mínimo questionável. Esse tema foi recentemente decidido pelo Supremo Tribunal Federal, que entendeu, por 9 votos a 2, que essa norma é conforme a constituição e que a Receita Federal pode acessar informações bancárias de contribuintes sem autorização judicial nos casos de apuração de fraudes fiscais.[39]

C. Acesso: Ainda que o Banco Central do Brasil

[38] Disponível em Normas SRF: http://normas.receita.fazenda.gov.br/sijut2consulta/consulta.action

[39] Ver http://agenciabrasil.ebc.com.br/economia/noticia/2016-02/stf-confirma-que-receita-pode-acessar-dados-bancarios-sem-autorizacao

tenha começado a projetar sistemas capazes de dar aos cidadãos acesso a partes de suas próprias informações financeiras pessoais (por exemplo, o sistema Registrato)[40], faltam instrumentos e determinações legais para garantir o acesso de indivíduos a informações pessoais em cada novo sistema concebido. Ao que parece, o acesso dos indivíduos não é a regra, e só deve ser implementado se houver uma justificativa específica. Além disso, os indivíduos que requerem acesso devem seguir procedimentos que não foram concebidos especificamente para acessar dados, o que pode ser demorado e fastidioso.

[40] http://www.bcb.gov.br/pt-br/sfn/registrato/Paginas/default.aspx

5/ Possível impacto da futura regulamentação

Considerando a legislação brasileira atual sobre proteção de dados, no momento não há nenhuma legislação aplicável que possa servir de base para o tratamento de dados. À luz do Anteprojeto de Lei sobre Proteção de Dados Pessoais, apresentado pelo Ministério da Justiça brasileiro em outubro de 2015, é razoável argumentar que os sistemas PNR e API poderiam ser classificados[41] na categoria do Art. 4º, inciso III:

Art. 4 – Esta lei não se aplica ao tratamento de dados:
...
III – realizado para fins exclusivos de segurança pública, defesa nacional, segurança do Estado, ou atividades de investigação e repressão de infrações penais.

O assunto está, portanto, fora do alcance da legislação proposta para proteção de dados, exceto por uma consideração que segue no mesmo artigo 4º, § 1º do anteprojeto:

§ 1º O tratamento de dados pessoais previsto no inciso III será regido por legislação específica, observados os princípios gerais de proteção e os direitos do titular previstos nesta Lei.

Tais "princípios" e "direitos do titular" são descritos posteriormente nas regulamentações propostas. Estes incluem diretos como o acesso à informação a ser exercido pelos cidadãos (Art. 8º do anteprojeto de lei), bem como outras disposições que devem fortalecer a posição dos cidadãos sem entrar em conflito com a finalidade dos sistemas.

Como observado anteriormente, o acesso individual aos próprios dados é concedido pela ação de Habeas Data, e seu processo é implementado por meio da Lei de Acesso à Informação (Lei 12.527 de 2011); entretanto, estes não são procedimentos adequados para os problemas atuais de proteção de dados (por exemplo, o primeiro só pode ser solicitado por um advogado e o segundo só pode ser usado perante órgãos públicos).

Outra possível influência da regulamentação proposta é a definição que dá a termos como "dados pessoais", "dados sensíveis" e "dados anonimizados", em seu artigo 5º, todos eles presentes nos sistemas analisados. Outro

[41] Available at Normas SRF: http://normas.receita.fazenda.gov.br/sijut2consulta/consulta.action

possível impacto existiria no caso de o anteprojeto criar uma autoridade para proteção de dados. Nessa situação, a autoridade poderia, de acordo com o mesmo artigo 4º da legislação proposta, emitir recomendações ou solicitar aos responsáveis uma Avaliação de Impacto na Privacidade:

§ 3º O órgão competente emitirá opiniões técnicas ou recomendações referentes às exceções previstas nos incisos II e III, bem como poderá solicitar aos responsáveis relatórios de impacto à privacidade.

[41] Available at Normas SRF: http://normas.receita.fazenda.gov.br/sijut2consulta/consulta.action

Conclusão:

Embora a coleta e o tratamento de dados pessoais por meio dos sistemas examinados possa ser associada com um interesse público relevante – e este é, com efeito, a justificativa para sua implementação –, o fato é que tal procedimento levanta várias questões concernentes à proteção de dados.

Em primeiro lugar, há alguns pontos em sua implementação nos quais a falta de transparência é concreta – por exemplo, na ausência de determinações legais com relação ao acesso dos indivíduos a seus próprios dados ou, o que é ainda mais preocupante, no contexto em que a SRF recorre às redes sociais, uma atividade que foi meramente mencionada, sem nenhuma alusão aos métodos e critérios aplicados. Além disso, essa prática indica que pode haver outras fontes do interesse da SRF que estejam sendo usadas ou que venham a ser usadas no futuro, o que requer mais transparência.

Deve-se considerar, também, a finalidade e a necessidade dos dados coletados com relação ao objetivo do tratamento de tais dados. Embora as razões de um interesse público relevante devam ser levadas em conta, pode haver questões concernentes ao uso secundário de dados – por exemplo, de dados que podem ser considerados confidenciais, ou mesmo de dados que podem ser considerados não necessários para os fins almejados.

Entretanto, a carência de uma lei geral de proteção de dados no Brasil faz que considerações como estas, com relação a transparência, finalidade e necessidade, não sejam tão óbvias, já que esses princípios ainda não estão arraigados no marco jurídico do país. Portanto, fica claro que a promulgação de um marco regulatório para a proteção de dados será benvinda a fim de forçar estes e outros sistemas a cumprirem regras claras no que concerne aos pontos mencionados e a outros direitos dos titulares dos dados, além de proporcionar confiança aos vários atores envolvidos.