CONTRIBUIÇÃO PARA O DEBATE SOBRE PO PROJETO DE LEI DE PROTEÇÃO DE DADOS

Mariana Cunha e Melo

Bacharel em Direito pela UERJ, Mestre em Direito pela New York University School of Law. Advogada associada no escritório de advocacia Barroso Fontelles Barcellos Mendonça & Associados. Advoga em Brasília.

I. UM CONCEITO ABRANGENTE DE DADO PESSOAL DESINCENTIVA PRÁTICAS QUE AUMENTAM A SEGURANÇA DE DADOS

Dispositivo do Projeto de Lei:

> "*Art. 5º Para os fins desta Lei, considera-se:*
> *I – dado pessoal: dado relacionado à pessoa natural* ***identificada ou identificável, inclusive a partir de números identificativos****, dados locacionais ou identificadores eletrônicos*".

Comentário:

O projeto de lei adota um conceito abrangente de dado pessoal, que engloba dados *identificados* e *identificáveis*. Isso quer dizer que a lei incidiria sobre dados que identificam claramente o usuário; os dados agregados por meio de identificadores numéricos (chamados identificadores anônimos persistentes), que permitem a formação de um perfil da navegação de um usuário sem que se identifique nominalmente o mesmo[1]; e os dados que, muito embora tenham sido desidentificados, retirando marcadores pessoais ou numéricos que poderiam indicar a origem do dado, são identificáveis quando cruzados com outras bases de dado[2].

[1] É o caso quando plataformas de serviço são capazes de identificar que a mesma pessoa que acessou determinado site ontem realizou uma compra hoje, criando um perfil de comportamento dessa pessoa, muito embora a plataforma de serviço não saiba quem ela é fora do mundo virtual.Ver: Solon Barocas and Helen Nissenbaum, "Big Data's End Run around Anonymity and Consent," in Julia Lane, Victoria Stodden, Stefan Bender, Helen Nissenbaum, eds. *Privacy, Big Data, and the Public Good Frameworks for Engagement* (Cambridge Univ. Press 2014), p. 51.

[2] Felix T. Wu, "Defining Privacy and Utility in Data Sets," 84 University of Colorado Law Review 1117, 1137-44 (2013). Sobre técnicas de desidentificação de dados, ver Opinion 05/2014 on Anonymisation

Este comentário é direcionado à inclusão da noção de dados identificáveis no conceito de dados pessoais. Há um interesse público no equilíbrio entre a desidentificação de dados e a manutenção de alguns marcadores de identificação. Quando detentoras de bases de dados pessoais agregam muitas informações sobre indivíduos, há um risco que essas bases sejam hackeadas e os dados sejam não-intencionalmente vazados. Idealmente, portanto, essas detentoras de bases de dado deveriam desidentificar os dados na medida do possível. Mas nem tudo pode ser desidentificado. Do outro lado da balança está a utilidade dos dados, que muitas vezes depende da manutenção de alguns marcadores de identificação[3].

Mesmo os dados desidentificados, porém, não estão, via de regra, inteiramente livres do risco de violação da privacidade de dados. Pesquisas recentes demonstraram que, dependendo da técnica que se use para desidentificar dados, é possível a re-identificação por meio do cruzamento de duas ou mais bases de dados distintas[4]. Por isso, a doutrina raramente fala em dados desidentificados, mas em dados identificáveis, uma vez que não são imediatamente direcionados a um indivíduo em particular, mas podem ser por meio de técnicas de re-identificação. O risco de re-identificação varia de caso a caso, mas essa possibilidade, sustentam alguns, sempre existiria[5].

Techniques of the Working Party On The Protection Of Individuals With Regard To The Processing Of Personal Data, disponível em <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf>.

[3] Um exemplo pode facilitar a compreensão do ponto. Dados relativos à saúde das pessoas são especialmente sensíveis. Por conta disso, alguém poderia sustentar que esses dados não poderiam ser retidos pelas prestadora de serviço ou que se deveria desidentifica-los inteiramente, para reduzir o risco de vazamento de dados pessoais. Algum nível de identificação desses dados, porém, é indispensáveis para o avanço da medicina. Para pesquisas em saúde, dados sobre a incidência de doenças (a Dengue, por exemplo), somente são úteis para desenvolvimento de políticas de controle da doença se conhecidas algumas características do paciente, como o gênero, a idade, a condição física, o endereço de residência, etc.. Impor a anonimização de todas as bases de dados, portanto, seria uma tolher o avanço de pesquisas que usam dados como utilidade.

[4] Felix T. Wu, "Defining Privacy and Utility in Data Sets," 84 University of Colorado Law Review 1117, 1137-44 (2013). Sobre técnicas de desidentificação de dados, ver Opinion 05/2014 on Anonymisation Techniques of the Working Party On The Protection Of Individuals With Regard To The Processing Of Personal Data, disponível em <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf>.

[5] Paul Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization.

Por conta desse risco de re-identificação dos dados desidentificados, a regulação europeia de proteção de dados passou a incluir a noção de dados identificáveis no conceito de dados pessoais, tradição a que adere o projeto de lei. **Essa postura, porém, cria mais riscos que proteções aos dados pessoais e deve ser retirada do projeto. Isso porque o projeto de lei, ao equiparar dados identificados a identificáveis, retira um incentivo relevante para que as detentoras de bases de dados desidentifiquem esses dados[6].** Considerando que as restrições legais seriam as mesmas para dados plenamente identificados e dados desidentificados mas identificáveis, não há por que as empresas detentoras de bases de dado investiriam na anonimização dos dados.

Nos EUA, a legislação de privacidade se restringe aos dados identificados, por duas razões, ao menos. A primeira razão é lógica. A legislação constritiva só tem razão de ser para proteger a privacidade dos indivíduos e, se o dado não oferece risco de identificação do mesmo, tampouco há razão para a proteção especial, a princípio. **A segunda razão é de ordem pragmática. Como estratégia de *compliance* com a legislação de privacidade, detentoras de dados têm o interesse de evitar que a legislação se aplique às suas bases de dado, para evitar as constrições. Para isso, eles desidentificam suas bases e, sem dados identificados, não incide a legislação específica.**

Para equilibrar o interesse na proteção de dados que são identificáveis e que portanto oferecem algum risco de re-identificação com o interesse em manter incentivos para a desidentificação mais segura possível, dois dos maiores doutrinadores em direito da proteção de dados no Estados Unidos formularam a seguinte proposta, intitulada "dados pessoais 2.0"[7]. O definição é tripartite. Além das tradicionais conceitos de dados identificados e não-identificáveis, criaram a noção de dados identificáveis como dados que estão sob um nível relevante de risco de re-identificação. Essa classificação tripartite traz duas inovações interessantes. A primeira é que não basta ser identificável em tese para ser enquadrado nesse conceito, mas representar um risco

[6] Sobre essa crítica ao modelo europeu, v.: Paul M Schwartz e Daniel J Solove, "The PII Problem: Privacy and a New Concept of Personally Identifiable Information," 86 N.Y.U. L. Rev. 1814, 1873-77 (2011); e Paul M Schwartz e Daniel J Solove "Reconciling Personal Information in the United States and European Union", 102 California Law Review 877 (2014).

[7] Paul M Schwartz and Daniel J Solove, "The PII Problem: Privacy and a New Concept of Personally Identifiable Information," 86 N.Y.U. L. Rev. 1814, 1877 (2011).

relevante de ser de fato re-identificado. A segunda é que, mesmo quando considerada a informação identificável, os autores sustentam que apenas uma parcela da proteção legal deve incidir. Afinal, quanto menor o risco de violação da privacidade de dados, menor deve ser a proteção a essas informações e consequentemente as amarras legais relativas a tais dados[8].

Dessa forma, se mantem o incentivo às detentoras de bases de dados para anonimizar seus dados da forma mais segura possível e se protege em certa medida os dados identificáveis.

II. O FOCO DA PRIVACIDADE DE DADOS NÃO DEVE SER NO CONSENTIMENTO PRÉVIO E DESTACADO DOS INDIVÍDUOS

Dispositivo do Projeto de Lei:

> "Art. 7º O tratamento de dados pessoais somente é permitido após o consentimento livre, expresso, específico e informado do titular, salvo o disposto no art. 11. (…)
> §3º O consentimento deverá ser fornecido por escrito ou por outro meio que o certifique.
> §4º O consentimento deverá ser fornecido de forma destacada das demais cláusulas contratuais.
> §5º O consentimento deverá se referir a finalidades determinadas, sendo nulas as autorizações genéricas para o tratamento de dados pessoais. (…)
> Art. 10º No momento do fornecimento do consentimento, o titular será informado de forma clara, adequada e ostensiva sobre os seguintes elementos:
> I – finalidade específica do tratamento;
> II – forma e duração do tratamento;
> III – identificação do responsável;
> IV – informações de contato do responsável;
> V – sujeitos ou categorias de sujeitos para os quais os dados podem ser comunicados, bem como âmbito de difusão;
> VI – responsabilidades dos agentes que realizarão o tratamento; e
> VII – direitos do titular, com menção explícita a:
> a) possibilidade de não fornecer o consentimento, com explicação sobre as consequências da negativa, observado o disposto no § 1º do art. 6º;
> b) possibilidade de acessar os dados, retificá-los ou revogar o consentimento, por procedimento gratuito e facilitado; e

[8] Paul M Schwartz and Daniel J Solove, "The PII Problem: Privacy and a New Concept of Personally Identifiable Information," 86 N.Y.U. L. Rev. 1814, 1877 (2011).

> c) possibilidade de denunciar ao órgão competente o descumprimento de disposições desta Lei.
> § 1º Considera-se nulo o consentimento caso as informações tenham conteúdo enganoso ou não tenham sido apresentadas de forma clara, adequada e ostensiva.
> § 2º Em caso de alteração de informação referida nos incisos I, II, III ou V do caput, o responsável deverá obter novo consentimento do titular, após destacar de forma específica o teor das alterações.
> § 3º Em caso de alteração de informação referida no inciso IV do caput, o responsável deverá comunicar ao titular as informações de contato atualizadas.
> § 4º Nas atividades que importem em coleta continuada de dados pessoais, o titular deverá ser informado regularmente sobre a continuidade, nos termos definidos pelo órgão competente".

Comentário:

A exigência de consentimento prévio do usuário para cada atividade de tratamento de dados é inócua e sua implementação é inviável. Em primeiro lugar, o Marco Civil exige a retenção de dados por provedores de aplicações e de conexão. Ou seja: esses provedores não têm escolha e são obrigados a coletar e armazenar dados pessoais de forma identificada. Consequência disso é que dizer que esses provedores necessitam do consentimento de seus usuários para coletar e armazenar dados é sem propósito. Não há essa escolha. A ordem de obtenção do consentimento dos usuários torna-se ainda mais curiosa considerando que o anteprojeto de lei veda que o consentimento seja condição para o fornecimento de serviço. O projeto é, portanto, incongruente com a obrigação de retenção de dados prevista no Marco Civil, que faz a exigência de consentimento inócua. A questão da retenção de dados será retomada adiante.

Em segundo lugar, é lugar comum na doutrina sobre privacidade de dados que a exigência de consentimento para tratamento de dados é vazia de significância prática. Diversos estudos demonstram que usuários não leem termos de serviço[9] — elas simplesmente clicam "eu aceito". Estudos sugerem o índice de leitores dessas políticas de privacidade não aumenta mesmo quando facilitado o acesso ao texto ou quando destacados temos mais relevantes para serem

[9] Sobre o tema, v. Yannis Bakos, Florencia Marotta-Wurgler, e David R. Trossen, Does Anyone Read the Fine Print? Consumer Attention to Standard-Form Contracts, The Journal of Legal Studies, Vol. 43, No. 1 (January 2014), pp. 1-35, disponível em: http://www.jstor.org/stable/10.1086/674424.

aceitos[10]. A doutrina destaca que, dada a quantidade de serviços online que coletam e utilizam os dados pessoais e que os usuários acessam todos os dias, caso alguém se disponibilizasse a ler todos os termos que outros aceitam sem ler, essa pessoa gastaria em média um mês por ano apenas lendo os termos[11]. O aumento de itens para clicar "eu aceito", portanto, não garante que usuários fiquem melhor informados — pelo contrário, reduz a probabilidade que os usuários leiam qualquer das cláusulas que apareçam para eles aceitarem.

Por fim, e em terceiro lugar, a exigência de consentimento específico para cada uso de dados é inteiramente incompatível com o conceito de big data. A ideia por trás da indústria do big data é que dados cotidianamente produzidos por usuários em todo o planeta podem ser utilizados para a pesquisa e o avanço da humanidade. Exemplo emblemático disso foi a descoberta do efeito colateral de dois medicamentos individualmente aprovados pela autoridade de controle dos EUA, mas que se ministrados em conjunto aumentavam o índice de glicose no sangue de pacientes diabéticos[12]. A descoberta foi feita por meio da análise da base de dados de busca do Bing (provedor de pesquisa da Microsoft). Pesquisadores cruzaram a lista de usuários que fizeram a pesquisa pelo nome de cada medicamento individualmente e sintomas característicos como "dor de cabeça" ou "fatiga" e concluíram pela presença de efeitos colaterais, salvando a vida de milhões de pacientes. **Seria inviável realizar tal pesquisa caso fosse necessário o consentimento de cada usuário do Bing individualmente e por escrito.**

O consentimento prévio é, portanto, tanto inócuo, porque os usuários comprovadamente não leem os termos antes de concordar com os mesmos, quanto inconvenientes, porque tolhem a funcionalidade das pesquisas com big data. Melhor solução seria manter a obrigação de disponibilização dos termos em local de fácil acesso desde antes da contratação do serviço e manter aberta a oportunidade para retirada do serviço a qualquer tempo, quando as circunstâncias do serviço não recomendarem outra solução.

---

[10] Yannis Bakos, Florencia Marotta-Wurgler, e David R. Trossen, Does Anyone Read the Fine Print? Consumer Attention to Standard-Form Contracts, The Journal of Legal Studies, Vol. 43, No. 1 (January 2014), pp. 1-35, disponível em: http://www.jstor.org/stable/10.1086/674424.

[11] Omer Tene and Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics.*

[12] Omer Tene and Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, p. 8.

III. RETENÇÃO DE DADOS COMO A VIOLAÇÃO MÁXIMA À PROTEÇÃO DE DADOS

O projeto de lei de proteção de dados pessoais deveria se valer do ambiente propício à defesa da privacidade de dados e revogar a obrigação de retenção de dados prevista no Marco Civil. A retenção de dados foi julgada inconstitucional em diversos países e pela União Europeia[13] e rejeitada no debate político de tantos outros[14].

IV. DEFESA PROCEDIMENTAL DOS DADOS PESSOAIS

A lei de proteção de dados também deveria impor restrições procedimentais às disposições do Marco Civil que previu a divulgação de dados pessoais às autoridade judiciais e do Ministério Público. Confira-se:

A. Need for minimal standards for disclosure

The absence of minimal standards for the disclosure of users' identifying information was one of the reasons the European Court of Justice struck down the Data Protection Directive. The Court argued that a criterion was necessary to restrain the cases where law enforcement agencies could require users' data. The possibility of disclosure should depend on the "seriousness of the interference with the fundamental rights in question" as well as the seriousness of the criminal activity it intended to prevent, detect or persecute. The ECJ held that the directive did not meet this requirement because it only mentioned the fight against "serious crimes" as a reason to disclose the personal data[15].

As to the Marco Civil and the Brazilian case law on disclosure of identifying information, both of them fail to provide any sort of objective standard for disclosure, allowing a high presumption against anonymity. Even worse, the Marco Civil foresees that specific statutes

---

[13] Ver: http://curia.europa.eu/jcms/upload/docs/application/pdf/2014-04/cp140054en.pdf.

[14] Para dar alguns exemplos: Argentina, Austria, Bulgaria, Chipre, República Tcheca, Alemanha, Romania, Slovenia e Paraguai.

[15] See Court of Justice of the European Union, Judgment in Joined Cases C-293/12 and C-594/12, Press Release n. 54/14, Published on 09/07/2014.

may authorize law enforcement authorities to demand access to personal data without judicial intervention[16].

In light to constitutional interests that favor anonymity, the Courts should narrow the scope of both the statute and the precedents. The application of the requirements for adequate and necessary restriction of free speech, privacy and data protection to further the right to reputation in this scenario means establishing a test for the disclosure that: (i) requires the requesting party to show high probability of success on the merits of the libel action against the anonymous individual – otherwise, the disclosure would not be adequate to protecting the honor of the requesting party; and (ii) requires a showing that the identification of the anonymous individual is indispensable to protect the honor, that is, the inexistence of a less restrictive means – such as the right to reply.

To identify the rigor applied such test, the first step is to separate cases in which the defendant is being sued in her speaker capacity and those in which the content is just evidence to a non-speech-related tort or crime. In the first case, clearly free speech interests, privacy and data protection are present, while in the second case at least data protection interests applies anyway. Once identified the interests in conflict with the prohibition of anonymity, the standard applicable to the balancing test should take into consideration the relevance of the interests in favor of anonymity. This means that, for instance, that whenever a political speech is under request for disclosure of its author, the test for revealing the information should be the highest one. If, however, the speech does not seem to have any particular content but pure fighting words, the standard should be the lowest.

Thus, depending on the level of scrutiny applied, Courts should authorize the disclosure only upon a finding of: (1) higher or lower likelihood of the commitment of an unlawful behavior by the user; (2) no alternative reasonable means for (a) finding the authorship of the crime or (b) seeking redress for violation of intimacy, honor or reputation[17]; and (3) "need for the discovery sought [that] outweighs' the free expression rights of the person wishing to

[16] Marco Civil, article 10, § 3rd: "The provision of the caput of Art. 10 does not prevent administrative authorities to have access to recorded data that informs personal qualification, affiliation and address, as provided by law.".

[17] See dissent and concurrence opinions' disclosure test in Branzburg v. Hayes, 408 U.S. 665 (1972).

remain anonymous. See, for example, *Doe v. 2theMart.com*, 140 F. Supp. 2d 1088, 1095 (W.D. Wash. 2001)"[18].

In determining the likelihood of success in the merits or a libel action, the Courts should consider: (i) the truthfulness of the fact, (ii) the legality of the means employed to obtaining the information; (iii) whether the offended party is a public or private person; (iv) the place where the fact occurred; (v) the nature of the fact; (vi) the public interest in disclosing the fact; (vii) the public interest in disclosing facts related to the conduct of public officials; (viii) preference for after-the-disclosure punishment over prior restraints.

B. Minimal due process standards: suing the messenger, not the speaker

Currently the law requires service providers to collect and retain users' personal information in all cases and to disclose it upon preliminary injunction[19]. This gives rise to another issue that does not concern the substantive law on the matter. It is rather a procedural issue.

It is common knowledge that in order to state a claim against someone, the plaintiff must show that the defendant offended a right held by the plaintiff. For instance, under Brazilian procedural law, anyone who is entitled to receive certain documents from a company due to a contractual or statutory right has a civil action to require the documents in Court. At the end of such litigation the judge either holds the plaintiff entitled to the documents and force the defendant to deliver them or holds the plaintiff does not enjoy such right and dismiss the claim. Either way, the loser pays all legal and statutory attorney[20] fees. Therein lies the problem.

---

[18] Electronic Frontier Foundation, Anonymity and Encryption (feb.2015), p. 18. Available at <https://www.eff.org/document/eff-comments-submitted-united-nations-special-rapporteur-promotion-and-protection-right>.

[19] In this particular, the case law is in accordance with the EFF's recommendations on anonymity and encryption. See Electronic Frontier Foundation, Anonymity and Encryption (feb.2015), p. 18. Available at <https://www.eff.org/document/eff-comments-submitted-united-nations-special-rapporteur-promotion-and-protection-right>: "Judicial systems, not extrajudicial decision-making processes, are best suited to balance citizens' right to anonymous expression with the need to provide a mechanism to redress wrongs. Therefore, it is imperative that the laws do not require or permit Internet intermediaries to reveal the identity of the users without judicial decision-making. But judicial systems can only function when a court has an opportunity to review the circumstances before the identity is revealed. Therefore, to protect citizens' fundamental rights of freedom of expression and privacy, Internet intermediaries should only disclose the identity of an anonymous or pseudonymous user of their platform or service upon receipt of a court order, granted after a process of judicial review.".

[20] Typically, it varies in a range of 10-20% of the economic expression of the cause.

When a private citizen wants to uncover the identity of anonymous users, the procedural tool she typically uses is the civil action to disclose documents, or even a generic civil action against the service provider. Three issues arise from this: a theoretical and two practical ones. The first theoretical issue is that the intermediary did not give cause to the action. It did nothing unlawful. Rather the opposite, the service provider could not disclose personal information without a court order. Thus, service providers are sued for complying with their legal duty of protecting its users' data in the absence of a preliminary injunction. As to the practical issues, the *first* is that, as a result of being the defendant in the lawsuit, service providers pay all legal fees whenever the Court holds in favor of the plaintiff[21] – again: for fulfilling their legal duty of retaining and securing users' data. The *second* practical issue is that it creates a due process issue for the user whose personal information is to be revealed. That is because the user will never be able to argue in favor of its own anonymity. In fact, many users could never know about the intention of unmasking their anonymity until it is too late and the damage to their constitutional rights is done.

To address these issues, the law must adopt a different approach as to the procedural mechanism to requiring the disclosure of anonymous users. Service providers should not be put in the position of being the only person capable of arguing in favor of free speech and data protection in a disclosure lawsuit. The procedure should ensure adversary hearing before granting the preliminary injunction. In this context, the service provider should be authorized to send a notice to the user whose identity is being sought. Moreover, the true defendant – the user – should have not only mere notice of the claim[22], but also a certain amount of time to reply and defend its own anonymity[23-24]. In this scenario, the service provider would receive the reply and

---

[21] Even worse, the latest development of the jurisprudence about intermediaries responsibilities allows plaintiffs to convert the request for users' data into a complaint for damages, in total disregard of the Brazilian civil procedure, which does not prescribe such conversion. See STJ, 11-26-2013, REsp 1398985/MG, J. NANCY ANDRIGH; and STJ, 03-10-2014, REsp 1417641/RJ, J. NANCY ANDRIGHI.

[22] See Dendrite Intern. v. Doe No. 3, 775 A.2d 756 (2001) and *Doe v. Cahill*, 884 A.2d 451 (Del. 2005).

[23] See e.g. Legal Protections for Anonymous Speech in Virginia, Digital Media Law Project (Jan. 8th, 2014), http://www.dmlp.org/legal-guide/legal-protections-anonymous-speech-virginia. See also EFF, Test for Unmasking Anonymous Speech, Internet Law Treatise. Available at http://ilt.eff.org/index.php/Speech:_Anonymity#Tests_for_Unmasking_Anonymous_Speakers; "As best practices, those third parties should: Make reasonable efforts to notify the person whose identity is sought; If possible, agree to a timetable for disclosure of the information to the party seeking it that provides a reasonable opportunity for the Internet user to file an objection with a court before disclosure; Forward the exact statements and material provided by the person seeking the identity, including information about the cause of action alleged in the lawsuit and the evidence provided by the

reproduce it integrally on the records so the court could analyze all arguments in favor and against unmasking the user. Finally, the procedure should also provide a solution so that service providers do not figure as defendant and do not be held accountable for plaintiff's legal fees.

identity-seeker to the court where provided to the service provider.". Electronic Frontier Foundation, *Anonymity and Encryption* (feb.2015), p. 18. Available at <https://www.eff.org/document/eff-comments-submitted-united-nations-special-rapporteur-promotion-and-protection-right>.

[24] See also http://mediadefence.org/sites/default/files/files/20140606%20Delfi%20intervention%20FINAL.pdf: "A claim can be filed against the intermediary to discover identifying details of the original publisher so that the claim can be made against the proper party. In such cases, courts typically require the claimant to make a prima facie showing demonstrating a likelihood of a valid claim and provide an opportunity for an anonymous commenter to respond. See Dendrite Int'l, Inc. v. John Doe No. 3, 775 A.2d 756 (N.J. Super. Ct. App. Div. 2001).".