

Dados “anônimos” como antítese de dados pessoais: o filtro da razoabilidade

Série: Impactos operacionais e normativos da LGPD

A LGPD não se aplica a dados anonimizados. Com a recorrente publicação de estudos demonstrando as possibilidades de reversão de técnicas de anonimização, faz-se necessário estabelecer critérios e pressupostos para determinar o que configuraria, de fato, um dado anonimizado.

O segundo tema da série é a razoabilidade como filtro à identificabilidade de dados.

A antítese do conceito de dado pessoal seria um dado anônimo, ou seja, aquele que é incapaz de revelar a identidade de uma pessoa. Diante do próprio significado do termo, anônimo seria aquele que não tem nome nem rosto.

Essa inaptidão pode ser fruto de um processo pelo qual é quebrado o vínculo entre o(s) dado(s) e seu(s) respectivo(s) titular(es), o que é chamado de anonimização. Esse processo pode se valer de diferentes técnicas que buscam eliminar tais elementos identificadores de uma base de dados, variando entre: **a)** supressão; **b)** generalização; **c)** randomização; e **d)** pseudoanonimização. Tendo em vista o escopo deste ensaio, tratarei apenas das duas primeiras com o objetivo de apontar as implicações normativas de uma eventual dicotomia entre dados anônimos (anonimizados) e dados pessoais.

Gerenciando a *identificabilidade* dos dados: limites e possibilidades das técnicas de anonimização

Tomemos como exemplo um banco de dados relacionais, i. e. aquele estruturado por tabelas, em que cada uma de suas colunas – que são chamadas de atributos – é a maneira pela qual os dados são organizados. É a correlação entre as colunas e as linhas dessa tabela que empresta valor (significado) aos dados, permitindo que deles seja extraído algo inteligível (informação):

A) Nome	B) CPF	C) CEP	D) Faixa etária	E) Classificação/ segmentação
1. Bruno dos Santos	123.456-77	04055-000	18	Jovem hipster
2. Bruno dos Santos	234.567-88	04055-111	17	Jovem poupador
3. Bruno dos Santos	345.678-99	04055-222	17	Jovem consumista
4. Bruna Souza	456.789-10	01201-000	65	Idoso com rentabilidade
5. Bruna Souza	567.891-01	04201-111	66	Idosa sem rentabilidade
6. Bruna Schonber	222.333.44-55	04201-222	70	Idosa com rentabilidade
7. Maria Silva	157.890.88-66	09201-000	40	Adulto desempregado
8. Maria Silva	666.666.66-66	09201-111	38	Adulto perfil executivo
9. Maria Sóstenes	987.354.22-99	09201-222	16	Jovem hipster

Tabela 1 – Base de dados relacionais

Neste caso, para a aplicação das técnicas de supressão e generalização, deve-se identificar quais elementos poderiam ser modificados – suprimidos ou generalizados – para que o seu *grau de identificabilidade* seja eliminado ou reduzido:

- a) **supressão do CPF:** por ser um identificador capaz de diferenciar até mesmo pessoas homônimas, sendo um identificador único; logo, a sua disponibilização, ainda que parcial – *e.g.*, cinco primeiros dígitos –, não seria prudente;

- b) generalização do nome completo:** constaria apenas o prenome, desde que fosse observado que os nomes da base de dados não são comuns. O objetivo é evitar que um nome possa ser atribuído a um indivíduo em específico;
- c) generalização da localização geográfica:** em vez de disponibilizar o número completo do CEP, seriam divulgados apenas os seus primeiros dígitos. Assim, haveria uma localização menos detalhada, a fim de quebrar o vínculo de identificação desta informação com um sujeito;
- d) generalização da idade:** em vez de divulgar a idade exata, seria divulgada a faixa etária para viabilizar a categorização dos indivíduos como jovens, adultos ou idosos (coluna “E”) e, por outro lado, inviabilizar a sua individualização, dado o universo de pessoas que se enquadram naquela mesma faixa etária.

A) Nome	B) CPF	C) CEP	D) Faixa etária	E) Classificação/ segmentação
1. Bruno dos Santos	123.456-77	04055-000	18>	Jovem hipster
2. Bruno dos Santos	234.567-88	04055-111	18>	Jovem poupador
3. Bruno dos Santos	345.678-99	04055-222	18>	Jovem consumista
4. Bruna Souza	456.789-10	01201-000	60<	Idoso com rentabilidade
5. Bruna Souza	567.891-01	04201-111	60<	Idosa sem rentabilidade
6. Bruna Schonber	222.333.44-55	04201-222	60<	Idosa com rentabilidade
7. Maria Silva	157.890.88-66	09201-000	18<	Adulto desempregado

8. Maria Silva	666.666.66-66	09201-111	18<	Adulto perfil executivo
9. Maria Sóstenes	987.354.22-99	09201-222	18>	Jovem hipster

Tabela 2 – Base de dados relacionais anonimizada

Com maior ou menor grau de intensidade – *e.g.*, supressão ou generalização – nota-se um método cujo mote é gerenciar circunstancialmente a *identificabilidade* de uma base de dados. As características de cada dado e a percepção de eles estarem inseridos em uma gama de informações devem orientar tal análise.

Por isso, não há um único método ou uma combinação perfeita *ex ante* para parametrizar o processo de anonimização, devendo-se analisar contextualmente como este deve ser empreendido para que os titulares dos dados anonimizados não sejam reidentificados, nem mesmo por quem procedeu à sua anonimização.

Amarrar o conceito teórico de dados anônimos a uma *análise contextual*, com os olhos voltados para a irreversibilidade do processo de anonimização, joga luz diretamente sobre o fator problemático dessa proposição: o seu caráter elusivo ou mesmo a sua impossibilidade teórica.

Torna-se cada vez mais recorrente a publicação de estudos que demonstram ser o processo de anonimização algo falível. A representação simbólica de que os vínculos de identificação de uma base de dados poderiam ser completamente eliminados, garantindo-se, com 100% (cem por cento) de eficiência, o anonimato das pessoas, é um mito. Sempre existirá a possibilidade de uma base de dados anonimizada ser agregada a outra para a sua reidentificação.

O conceito expansionista de dado pessoal e o filtro da razoabilidade

Por essa lógica, qualquer dado pessoal anonimizado detém o *risco inerente* de se transmutar em um dado pessoal. A agregação de diversos “pedaços” de informação (dados) pode revelar (identificar) a imagem (sujeito) do quebra-cabeça, a qual era até então desfigurada (anônimo) – o chamado efeito mosaico.

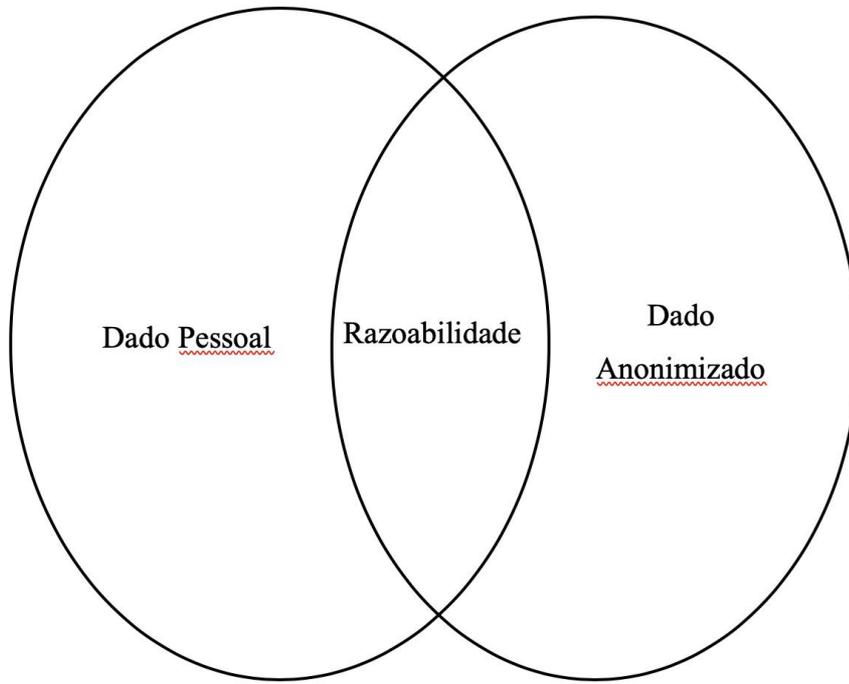
Por isso, em princípio, eventual dicotomia entre dados pessoais e dados anônimos só guardaria coerência junto ao conceito reducionista de dados pessoais. Isso porque dados anônimos não são dados relacionados a uma pessoa identificada, demandando a reversão do processo de anonimização para se chegar aos respectivos titulares, sendo a sua *identificabilidade* remota (identificável) e não imediata (identificada).

Dessa forma, leis que adotam o conceito expansionista de dados pessoais e, ao mesmo tempo, estabelecem uma dicotomia deste com dados anônimos correriam o risco de ser tautológicas. Isso porque haveria uma *redundância normativa*, já que dados anônimos seriam, em última análise, potencial e provavelmente, dados relacionados a uma pessoa identificável.

Para não gerar tal incoerência, a única saída foi a adoção de um “filtro” que delimitasse a *elasticidade* do conceito expansionista – neste caso, o termo identificável –, sob pena de a fronteira entre dados pessoais e dados anônimos ser sempre transponível.

E, nesse sentido, o direito comunitário europeu e a LGPD valeram-se do critério da razoabilidade para delimitar o espectro do conceito expansionista de dados pessoais. Não basta a mera possibilidade de que um dado seja atrelado a uma pessoa para atrair o termo identificável. Essa vinculação deve ser objeto de um “esforço razoável”, sendo esse o perímetro de elasticidade do conceito de dado pessoal como aquele relacionado a uma pessoa identificável.

A contrario sensu, se para a correlação entre um dado e uma pessoa demanda-se um esforço fora do razoável, não há que se falar em dados pessoais. Nessa situação, o dado é considerado como anônimo, uma vez que o “filtro da razoabilidade” barra o seu enquadramento como aquele relacionado a uma pessoa identificável. No próximo artigo da série, nós iremos traçar quais são os critérios significar o termo razoabilidade.



Com isso, há coerência em se estabelecer conceitos diferentes para tais espécies de dados, sobretudo sob o ponto de vista de uma dicotomia mutuamente excludente entre eles, que é delimitada pelo fator da razoabilidade. Do contrário, repita-se, haveria uma redundância normativa, na medida em que dados anônimos – sem o critério da razoabilidade – seriam sempre enquadrados dentro do conceito de dado pessoal, como aquele relacionado a uma pessoa identificável.

Bruno Ricardo Bioni. Doutorando em Direito Comercial e Mestre em Direito Civil pela Faculdade de Direito da USP. Trainee do European Data Protection Board e do Departamento de Proteção de Dados do Conselho da Europa. Fundador-professor do Data





Privacy Brasil e consultor na área de direito e tecnologia com ênfase em proteção de dados pessoais (www.brunobioni.com.br).

