

## Ensaio

# Dados pessoais e políticas públicas: O que revelar e o que preservar?

## Personal data and public policies: What to disclose and what to preserve?

Fernanda Teixeira Reis<sup>1\*</sup> , Denise do Carmo Direito<sup>2</sup> 

<sup>1</sup>Instituto Brasileiro de Geografia e Estatística, Escola Nacional de Ciências Estatísticas, Rio de Janeiro, RJ, Brasil

<sup>2</sup>Instituto de Pesquisa Econômica Aplicada, Brasília, DF, Brasil

Fernanda Teixeira Reis, branca, doutoranda da ENCE/IBGE, servidora licenciada do FNDE/MEC, editora associada da Revista Brasileira de Avaliação (RBMA)

Denise do Carmo Direito, branca, doutora em Ciência Política pela Universidade de Brasília, estágio doutoral na Universidade de Versailles Saint-Quentin-en-Yvelines (UVSQ - França), especialista em Políticas Públicas e Gestão Governamental (EPPGG), atuando como pesquisadora no IPEA.

**COMO CITAR:** Reis, Fernanda Teixeira, & Direito, Denise do Carmo. (2023). Dados pessoais e políticas públicas: O que revelar e o que preservar? *Revista Brasileira de Avaliação*, 12(1), e121623. <https://doi.org/10.4322/rbaval202312016>

### Resumo

Velocidade de transmissão e capacidade de armazenamento estão mudando a relação entre pessoas, mercado e Estado. Quando o Estado é o foco das análises, é importante tanto proteger os cidadãos quanto reproduzir como os dados são manipulados pela administração pública. O presente trabalho busca refletir sobre quais problemas encontrados para acessar os dados e caminhos que podem e devem ser percorridos para minimizar perda da privacidade e maximizar o uso de dados para a produção de evidências em políticas públicas. A promulgação da Lei Geral de Proteção de Dados (LGPD) é fronteira temporal de análise. Para as autoras, os dados pessoais devem ser tratados como um fato dinâmico e notadamente não punitivo. O Estado tem a obrigação de tratar e identificar os riscos potenciais de seu uso. Não pode, porém, adotar alternativas simples de restrição à informação ou optar pela disseminando apenas de forma agregada, opções que impossibilitam análises adequadas.

**Palavras-chave:** Estatística públicas. Evidência em políticas públicas. Avaliação e monitoramento. Lei Geral de Proteção de Dados (LGPD).

### Abstract

Transmission speed and storage capacity are changing the relationship between people, market and State. When the State is the focus of the analysis, it is important to protect citizens in the same way that it is important to reproduce how data is manipulated by the public administration. The present work seeks to contribute by reflecting on what are the current problems in the trend of access to public data and what are the options to minimize privacy problems and maximize access to public data used to produce evidence in public policies. For the authors, personal data should be treated as a dynamic and notably non-punitive fact. The State has the obligation to deal with and identify the potential risks of its use. The State cannot adopt simple restriction on access to information, nor disclose only aggregated data, because in both situations complex and important analyzes cannot be carried out.

**Keywords:** Official statistics. Evidence in public policy. Monitoring and evaluation (M&E). Lei Geral de Proteção de Dados (LGPD).

A RBAVAL apoia os esforços relativos à visibilidade dos autores negros na produção científica. Assim, nossas publicações solicitam a autodeclaração de cor/etnia dos autores dos textos para tornar visível tal informação nos artigos.

**Recebido:** Março 01, 2023

**Aceito:** Junho 16, 2023

**\*Autor correspondente:**

Fernanda Teixeira Reis

E-mail: teixeira.f@gmail.com



Este é um artigo publicado em acesso aberto (Open Access) sob a licença Creative Commons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

## Introdução

O mundo atual é mediado por plataformas, apps, sistemas e dados em profusão, em que informações são coletadas sistematicamente por empresas e órgãos públicos, a maior parte delas fornecidas espontaneamente por usuários das redes sociais ou para requerimento de bens e serviços. Uma simples pesquisa sobre o caminho mais perto, ou com menos trânsito, gera um extenso banco de informações que poderá ser utilizado para atividades diversas àquele fim, muitas delas monetizadas. Esse mundo super conectado, que teve início há pouco mais de 20 anos<sup>1</sup>, foi capaz de transformar a forma como ocorrem as relações entre as pessoas, o mercado e o Estado.

Na relação estado-sociedade o uso intensivo de dados, principalmente os pessoais e sensíveis, acarretaram tensões entre o direito à privacidade e a necessidade dessas informações para atender adequadamente à população: conceder benefícios de forma ágil, melhorar processos, otimizar políticas públicas e promover a transparência da atividade governamental e aplicação do orçamento público. De igual forma, ao mesmo tempo que é necessário resguardar o cidadão é imperativo dar transparência a forma como os dados são usados e manipulados pela administração pública, visto que muitas vezes são empregados recursos de tecnologias de Ciência de Dados, a exemplo de algoritmos, aprendizado de máquina (*machine learning*) e inteligência artificial (IA), processos complexos e de baixa transparência e de difícil compreensão para o conjunto da sociedade.

A partir desse empate entre dados sensíveis, privacidade e políticas públicas é que buscamos contribuir com a reflexão sobre quais os problemas atualmente encontrados para acessar os dados e quais os caminhos podem e devem ser percorridos para minimizar a perda da privacidade, mas maximizar o uso de dados para a produção de evidências em todas as fases das políticas públicas. Relembrando que estas são fundamentais para o avanço do país, enquanto fatores relevantes para a redução dos alarmantes índices de pobreza, diminuição das desigualdades sociais e contribuição ao crescimento econômico.

A fim de atingir esses objetivos, o presente artigo divide-se em quatro seções, contando com esta introdução. A seção 2 faz-se uma avaliação sobre o *locus* que os dados ocupam nas diferentes etapas do ciclo das políticas públicas, utilizando-se exemplos concretos em políticas sociais e educacionais e os limites e possibilidades identificados nesses diferentes processos. A seção 3 debate a dialética entre privacidade e proteção, tendo como ponto de partida a Lei nº 12.709, de 14 de agosto de 2018, mais conhecida com Lei Geral de Proteção de Dados (LGPD) e analisando-se como o uso de tecnologias sofisticadas, como anonimização e pseudoanonimização, podem compatibilizar essas duas pontes. Por fim, à título de conclusão, ressalta-se a necessidade do uso de dados e tecnologias em todos os ciclo das políticas públicas, inclusive aquele que permite que a sociedade conheça e avalie a ação pública.

## Uso de dados em políticas públicas

Estudar políticas públicas é deparar-se com um campo dispare, em que diversas teorias, métodos e modelagens concorrem e se completam. Originalmente, era um ramo comum à Ciência Política, porém, ele vem agregando contribuições de diversas áreas de conhecimento, como Administração Pública, Sociologia, Econometria, Estatística Social e Ciência de Dados. A Econometria, por exemplo, foi a primeira área de conhecimento a utilizar concretamente técnicas quantitativas para mensurar o impacto da ação do Estado na sociedade (Souza, 2005). Por sua vez, a estatística pública oferece subsídios informacionais à tomada de decisão nos diversos níveis institucionais e podem contribuir ao contínuo aprimoramento da política. A consolidação do uso de indicadores, por exemplo, está ligada à ascensão das atividades de planejamento do setor público ao longo do século XX (Bauer, 1967). Por sua vez, com a ampliação do poderio dos computadores, a Ciência de Dados vem utilizando-se de grandes registros administrativos, webscraping e outros aparatos de Big Data para aplicar algoritmos para fazer previsões de orçamento, seleção de público-alvo de programas públicos e avaliações de cenários.

<sup>1</sup> Consideremos aqui o marcador temporal 2000, quando a terceira geração de redes móveis, popularmente conhecidas por 3G, é implementada comercialmente no Japão, causando uma revolução nos dados móveis e na telecomunicação. (Dulcey, 2021).



No caso brasileiro, o amplo uso de indicadores públicos se iniciou a partir da década de 1980 com o emprego de indicadores de processos almejando maior controle de recursos escassos do Estado (Scaratti & Calvo, 2012). A década de 90 presenciou a crescente utilização de indicadores sintéticos com o objetivo de apreensão da realidade por meio de “uma única medida, resultante da combinação de múltiplas medições das suas dimensões analíticas quantificáveis” (Scandar Neto et al., 2008, p. 1). Estatísticas Públicas viabilizam iniciativas de monitoramento e avaliação, proporcionando evidências à ação governamental para melhoria do desempenho de programas governamentais, como também favorecem o exercício do controle social e reconhecem iniciativas exitosas na gestão das políticas públicas.

No que tange as políticas sociais, educacionais e de saúde, a consolidação do processo de descentralização e a ampliação de demandas locais pela formulação e implementação destas políticas fez com que “a geração e disseminação periódica de estatísticas, indicadores e índices municipais passem a assumir um caráter ainda mais relevante no processo de planejamento, sobretudo para subsidiar o processo de implantação, monitoramento e avaliação das políticas públicas” (Guimarães & Jannuzzi, 2004, p. 1). Esses autores defendem que o principal objetivo das estatísticas públicas é compreender a realidade a fim de garantir melhor gestão de recursos e programas sociais. Tratando-se de um país de tamanho continental, com mais de cinco mil unidades municipais autônomas, com diversas realidades locais, dados consolidados, municipalizados e disseminados é uma das peças-chaves para a boa execução das políticas públicas.

Seguindo essa linha, emerge no Brasil diversas ações e incentivos à prática de Políticas Públicas Baseadas em Evidências (PPBEs). PPBEs tradicionalmente se consolidam em contexto de racionalização de processos decisórios, no qual técnicas empíricas instrumentalizam análises racionais das políticas em questão. E dados são centrais para essas atividades, seja para instrumentalizar campos amostrais a *surveys*, ou mesmo para atribuir covariáveis a modelos estatístico de análises para prospectar correlações num contexto de complexidade do processo social, por exemplo. (Koga et al., 2022). Soma-se a esse cenário o fato de que na última década, com a ampliação do poderio computacional, vimos o crescimento do ramo da Ciência de Dados na área de públicas, a exemplo do uso de algoritmos e redes neurais para automatização de soluções e previsões de público-alvo.

O uso de dados em todo o ciclo das políticas públicas – da definição da agenda à implementação – é cada vez mais intensivo. Para fins analíticos dividiu-se esses usos em em três macroprocessos: (i) prospecção, avaliações e estudos; (ii) concessão de benefícios e (iii) em processos de auditoria e fiscalização. Sendo que em cada um foram analisados casos concretos e as eventuais dificuldades enfrentada para acesso ou os eventuais impactos não esperados gerados pelo uso das tecnologias e estatísticas.

#### i. Prospecção, avaliações e estudos

Só é possível dimensionar o atendimento do SUS – Sistema Único da Saúde ou do SUAS – Sistema Único da Assistência Social a partir de dados como, tipo de atendimento prestado, doenças ou vulnerabilidades mais prevalentes no território, padrão alimentar e nutricional dos habitantes, entre outras. São informações como essas que permitem a melhoria da estrutura de atendimento e a prestação de serviço à população. Outro exemplo, é a projeção do alunado possibilitado pelo Censo da Educação Básica (INEP/MEC), que se caracteriza por ser a linha mestra na condução dos principais programas educacionais brasileiros, a exemplo do Programa Nacional da Alimentação Adequada (PNAE), Programa Nacional do Livro Didático (PNLD) e Programa Nacional do Transporte Escolar (PNATE).

Os limites orçamentários e a crescente demanda social por bem-estar exigem políticas públicas cada vez mais eficazes, ou seja, capazes de cumprir os seus objetivos dispendendo o menor volume de recursos possíveis. Os dados e as informações geradas a partir de atividades avaliativas viabilizam esse permanente aprimoramento. Além do uso interno feito pelos órgãos e funcionários, no sentido de definir público-alvo, melhores desenhos de política e entre outras ações, é importante incentivar e dar as condições necessárias para que agentes externos, como universidades, pesquisadores e organizações da sociedade civil, também façam suas análises e, eventualmente, questionem a ação pública, contribuindo inclusive com o seu aperfeiçoamento.



Exemplo recente de como mudanças governamentais podem impedir ou limitar estudos e avaliações feitas por especialistas, Universidades e até por organizações da sociedade civil foi a limitação de divulgação dos microdados do Censo Escolar feita pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP/MEC) em 2021<sup>2</sup>. Naquele momento, somente dados desagregados de questionário “Escola” e parcialmente dos demais questionários passaram a ser disponibilizados. Limite esse que foi imposto inclusive a servidores de órgãos que necessitavam de microdados para planejar, monitorar e avaliar repasse dos programas educacionais federais, como os educacionais citados acima.

A LGDP foi utilizada como justificativa para a redução da disponibilização dos dados, como forma de evitar a identificação de pessoas. No entanto conforme asseveraram várias entidades e movimentos de Educação a própria LGPD (art. 7º, incisos II e III) assegura que “a administração pública pode realizar o tratamento de dados pessoais necessários ao cumprimento de obrigação legal e/ou execução de políticas públicas, sem que para isso seja necessário o prévio consentimento da/o titular destes dados” (Brasil, 2022). Além disso, os dados eram disponibilizados de forma anonimizada, ou seja, com um código que permitia unicizar a informação, mas sem a identificação do nome e outros dados pessoais das pessoas. Também dados de gestores educacionais, a exemplo se são concursados ou quais seu nível educacional, não se encontram mais disponíveis via microdados.

Ao limitar a divulgação desses microdados, impede-se ou limita-se a produção de indicadores. Por exemplo, inviabiliza-se a obtenção da taxa de distorção idade-série, por cor/raça ou faixa etária; taxa de estudantes com acesso a professores com nível superior; avaliação sobre o número médio de horas-aula por dia, impacto da gestão escolar à execução de programas, entre muitos outros cálculos e análises que só são possíveis com a utilização dos microdados. Também há um apagão da política educacional municipal, o que se configura enorme prejuízo não só a prestação de assistência técnica e monitoramento pelo órgão federal, como também por impossibilitar atividades de planejamento e monitoramento pelas prefeituras e seus órgãos executores.

É fato que o INEP/MEC permite acessar os microdados em sala segura, ou seja, um pesquisador(a) poderia analisá-los via agendamento prévio e deslocamento até Brasília-DF. É bastante evidente os limites que esse modelo impõe a todos que necessitam acesso aos dados. Exemplos simples, como estudantes de doutorado fora do contexto da capital federal, muito provavelmente não teria capacidade financeira de custear idas à Brasília. De igual forma, um gestor cidade pequena distante da capital também teria dificuldades de produzir análises e fazer o planejamento de programas e políticas educacionais do município. Além disso, as novas formas de abordagens de avaliações de políticas e programas públicos, com o uso de Ciência de Dados e seus modelos complexos de aprendizado de máquina, por exemplo, não estão disponíveis para serem utilizados nesse ambiente, inviabilizando esse tipo de análise.

## ii. Concessão de benefícios

A dimensão de dar acesso a dados pessoais, adotando os cuidados necessários para preservar individualidades, é evidente na dimensão do aperfeiçoamento das políticas públicas. Mas há uma segunda dimensão, bastante analisada em literatura específica (Filgueiras, 2023), mas ainda pouco debatida no âmbito da gestão. Trata-se da necessidade de dar transparências e construir mecanismos para que a forma e as práticas tecnológicas usadas dentro do Estado para conceder benefícios e avaliação das informações prestadas se tornem mais transparente.

A adoção de técnicas de ciências de dados tem sido adotada, principalmente, quando vem ao encontro de reduzir ou simplificar os processos internos do Estado. O uso, por exemplo, de algoritmos de aprendizado de máquina foi disseminado com o objetivo de analisar as solicitações de benefícios previdenciário. O INSS adotou esse tipo de solução e há indícios bastante evidentes da necessidade de seu aperfeiçoado. Conforme notícias da imprensa (Jornal Extra, 2022) e estudos realizados (Paiva & Pinheiro, 2021) o uso de ciência de dados, como algoritmos de classificação e seleção de beneficiários aptos à concessão de benefícios tem acarretado rejeição em massa, aparentemente, por sua aplicação de forma irregular.

<sup>2</sup> Brasil (2021).



Explicitar como essa análise vem sendo feita, a partir da divulgação dos parâmetros e dos algoritmos adotados é o primeiro passo para que solicitantes, estudiosos e interessados possam questionar eventuais erros nessas definições. De igual forma, a adoção da ciência de dados não pode ser o caminho para o insulamento burocrático. É urgente que esses processos venham acompanhados de mecanismos e rotinas claros para a reconsideração das análises feitas, incluindo-se aí a eventualmente necessidade de intervenção humana.

Esse último ponto indica que também para a necessária da construção de interfaces socio-estatais cada vez mais amigáveis para o cidadão. Sendo que essa definição deve levar em conta sempre a realidade brasileira, que é marcada pela desigualdade e pela existência de um significativo contingente de pessoas que não possuem acesso cotidiano aos canais de comunicação mediados pela internet, devendo ser considerado o estabelecimento de canais presenciais de atendimento a esse conjunto de pessoas.

### iii. Processos de auditoria e fiscalização

Outro uso frequente tem sido o cruzamento de informações de inúmeras bases governamentais para auditar e qualificar a concessão de benefícios. Trata-se de iniciativas com o objetivo de combater eventuais fraudes e concessões indevidas. Esse tem sido amplamente utilizada na concessão dos benefícios socioassistenciais, por exemplo. Tanto o programa Bolsa Família (Auxílio Emergencial/Auxílio Brasil) como o Programa de Benefício de Prestação Continuada (BPC) tem sido alvo de um crescente cruzamento de bases para auditoria, tendo o Cadastro Único para Programas Sociais, com dados autodeclarados pelos solicitantes de benefícios sociais, como rol inicial de beneficiários e para a concessão. São feitas, por exemplo, checagem com um número crescente de outros registros administrativos, como Cadastro Nacional de Informações Sociais (CNIS)<sup>3</sup>, para avaliação de correção das informações prestadas pelos requerentes (Banco Mundial, 2021).

Cabe ressaltar que esse processo apresenta problemas, já que as bases utilizadas muitas vezes são anteriores às informações prestadas ao Cadastro Único. Assim, o cidadão é obrigado a ir várias vezes ao local de cadastramento, como os Centros de Referência da Assistência Social (CRAS) prestar esclarecimentos, até conseguir ter autorizado os benefícios solicitados (BOLETIM IPEA 29). E embora essa pareça ser uma atividade simples, não o é: muitas vezes esses CRAS são distantes da moradia do cidadão, que nem sempre possui facilidade de contato ou até mesmo recurso financeiro, ou disponibilidade físico-temporal a exemplo de uma mãe com filhos pequenos, para pegar uma condução e se dirigir à entidade pública para corrigir esse registro.

Por motivos evidentes, as práticas administrativas de controle reverberam de forma positiva na sociedade. No entanto, a luta contra a fraude, quando não bem dosada, deixa de ser um processo de verificação administrativa e se transforma em um mecanismo central de cerceamento ao acesso a determinadas políticas públicas.

A adoção de batimentos com número cada vez maior de bases, poderia se transformar em uma “espiral rigorosista” (Dubois, 2021, p. 57), que não para de girar, apertando cada vez mais, sem que haja contraponto ao processo. Se considerarmos que esses controles são feitos prioritariamente com políticas destinadas aos mais vulneráveis, os aparentes indícios de inconsistência poderiam estar obliterando, na verdade, a dificuldade dessa população em encontrar meios não onerosos – recursos financeiros, de tempo, de compreensão – para se contrapor e esclarecer a avaliação indevida.

Nesse caso, a solução inicial e mínima para o conflito estaria na adoção e divulgação de processos transparentes sobre como os dados são tratados, além da adoção de interfaces socioestatais de fácil acesso para a revisão de exclusões indevidas. Dar transparência a forma como o Estado trata e manipula seus dados nos seus processos internos.

<sup>3</sup> Conforme relatório do Banco Mundial (2021), algumas bases utilizadas nesse processo de auditoria são: Imposto de Renda Pessoa Física (IRPF/RFB), o Sistema Integrado de Administração de Pessoal (SIAPE), a Folha de pagamentos de benefícios assistenciais e previdenciários do INSS (Maciça), a Relação Anual de Informações Sociais (RAIS), o Cadastro Geral de Empregados e Desempregados (CAGED)<sup>1</sup>, a Guia de Recolhimento do Fundo de Garantia do Tempo de Serviço e Informações à Previdência Social (GFIP) e o Sistema Integrado de Administração Financeira (SIAFI).



As pessoas deveriam de antemão saber o que vai ser feito com seus dados tanto em termos de repasse para outros órgãos, mas também do tratamento que sofrem. Passo fundamental é divulgar os microdados anonimizados e/ou pseudoanonimizados, e suas sintaxes de análises, para que a academia e sociedade civil as repliquem, encontrando possíveis falhas ou até mesmo proponham melhorias a seu uso. Algoritmos não podem suplantam agendas e decisões no ciclo da política pública. Tomadas de decisões via ciência de dados não podem ser caixas pretas, em que possíveis vieses criam mais desigualdades, por exemplo, em busca de um ótimo fiscal. São excelentes ferramentas que podem, e devem, ser debatidas, ampliadas, monitoradas e fiscalizadas.

### Dados: privacidade e proteção

Os exemplos acima confirmam que o Estado possui grande quantidade de dados que a depender do seu uso e divulgação podem causar severos danos aos cidadãos. São dados como saúde física e mental, escolaridade, trabalho e remuneração, entre outras. Por isso, a preocupação com a preservação da privacidade é permanente, lembrando que o atual mundo mediado pela tecnologia deixou a todos expostos a uma opinião pública alargada. No sentido de que dados do cidadão poderiam expô-lo ao escrutínio de uma sociedade amorfa e que interpreta erroneamente aquilo que foi revelado, causando prejuízos reais aos indivíduos ou conjunto deles.

É necessário preservar esses vários dados, dificultando o seu acesso por pessoas e sistemas não autorizados. No entanto, quando se pensa em políticas públicas, como os casos acima, fica claro que as tecnologias e os dados são ferramentas necessárias para o planejamento, a implementação, o monitoramento dessas políticas, bem como para a melhoria das rotinas de concessão de benefícios e serviços que beneficiem os cidadãos, fazendo com que palavras como IA – inteligência artificial – algoritmos, dataficação<sup>4</sup>, entre outras, invadissem o cotidiano da ação pública.

A privacidade ganhou a cena para garantir segurança de informações e proteger o sigilo de dados sensíveis de indivíduos na era da informação com o surgimento na Europa da *General Data Protection Regulation (GDPR)*, em 2016. No Brasil, ganhou um aliado com a promulgação da Lei Geral de Proteção de Dados (LGPD) instituída em 2018, a qual é um marco nacional na busca do uso dos dados pessoais de forma responsável. Sua promulgação limitaram a atuação governamental e apontou a necessidade de que as pessoas estejam informadas do que será feito com seus dados. Dessa forma, o limite de todo o processo de acesso a dados sensíveis pelo poder público está na compreensão de quais dados devem ser acessados considerando, principalmente, os resultados esperados.

Conforme prevê a LGPD, é permitida a utilização de dados sensíveis desde que o cidadão autorize o seu uso e manipulação. Como o caso INEP/MEC acima analisado esses processos encontram-se em fase amadurecimento, quando os limites do que pode e como deve ser feito ainda não estão completamente claros.

De um lado os dados, bem como as informações e análises geradas a partir deles, são centrais para que se dimensione políticas, orçamentos, público-alvo, formas de atendimento, desenho de políticas, resultados esperados, entre outras dimensões. Por outro lado, sua publicidade indevida pode causar severos danos a sociedade ao expor informações que de caráter pessoal e que violem a privacidade de cidadãos. A partir desse conflito é que se estabeleceu a necessidade de revelar e manipular dados, mas preservando identidades.

A decisão sobre o que revelar passa, inicialmente, por compreender aquilo que é de interesse do cidadão em um caso específico e da sociedade no genérico. O parâmetro deve passar pela avaliação de quais dados devem ser acessados e por que atores de forma a beneficiar a oferta e adequação de políticas públicas. Porém, há que se estabelecer amplo debate sobre essa ação. Por exemplo, para fazer busca ativa dos alunos que estão fora da escola, ou na iminência de abandonar, se faz importante ter uma chave que promova o acompanhamento desses alunos, a fim do Estado, especialmente na esfera municipal, adequar soluções a sua permanência.

<sup>4</sup> O termo “datafication”, traduzido para o português como dataficação, foi inicialmente cunhado Mayer-Schönberger & Cukier (2013) e se refere às formas de transformação de ações em dados quantificáveis, permitindo amplo rastreamento e análises preditivas.



Assim, o caminho do equilíbrio é trilhar dois caminhos. Disponibilizar o maior número de informações possíveis, sem expor identidades, seria o ótimo pretendido.

Nessa linha, há primeiro que se entender a diferença entre pessoa natural identificada ou identificável presente no Inciso primeiro, Art. 5º da LGPD<sup>5</sup>, endossado também pela Lei 12.527/2011 (Lei de Acesso à Informação- LAI<sup>6</sup>), em seu inciso IV, do Artigo 4º. Pessoa natural identificada está relacionada a uma informação direta ou indireta da mesma: nome, sobrenome, endereço, telefone, nome da mãe, data de nascimento de um registro administrativo, por exemplo. Por sua vez, pessoa natural identificável é quando dados de uma ou mais bases possibilitam a identificação de uma pessoa. Por exemplo, um conjunto de informações quando agregada pode viabilizar a identificação inequívoca de da identidade. Por exemplo hipoteticamente podemos imaginar que o seguinte conjunto de informações, como criança indígena, com baixa visão, estudante da única escola de educação infantil de algum pequeno município e beneficiária do BPC poderá tornar a identificação dessa criança de forma confiável para os moradores desse pequeno município. Agora some-se a esses dados todo o perfil da família, dados de renda, de saúde e tantas outras informações nas bases governamentais.

Nesse mesmo artigo, no inciso terceiro, a LGPD define o remédio para proteger essas pessoas: anonimização dos dados identificáveis. Por dados anonimizados, a LGPD define: “dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento”. Devem ser aplicadas técnicas que tornem a pessoa natural identificável.

Porém, a anonimização pode não ser a melhor técnica no caso de bases governamentais em tempos de grande poderio computacional, visto que o tratamento de dados de duas ou mais bases governamentais pode tornar a pessoa identificável. Assim, ascende em estudos de ciência de dados a pseudonimização, com técnica mais recentes a serem aplicadas em base de dados de pessoas naturais. A pseudonimização está relacionada a técnicas que enxergam dados de forma não binária, ou seja, como anônimo ou não anônimos, mas em um contínuo entre o dado anônimo e pessoal (EASAYS, 2015). O Grupo de Trabalho de Proteção de Dados do Artigo 29,<sup>7</sup> define o processo de pseudonimizar a substituição de atributos (com alta possibilidade de ser único) em um registro por outro. Melhor explicando, trata-se de um processo de disfarce (disguising) de identidade. Por exemplo, a criança indígena que exemplificamos acima seria disfarçada em uma escola com características semelhantes de outro município análogo.

A pseudonimização opera de maneira que as informações não podem ser conectadas a um titular de dados específico sem que se recorra a informações suplementares, desde que estas sejam mantidas separadamente, empregadas medidas organizativas e de segurança. Em pesquisas médicas, desde a década de 2000 já se estuda a implementação dessas técnicas a fim de se alcançar o respeito à privacidade dos sujeitos da pesquisa e o tratamento de dados de modo útil à investigação científica. Para estudos de radiologia clínica, a título de exemplo, a pseudonimização é importante para identificar dados de certo paciente de forma consistente com o passar do tempo – o que a anonimização (Machado & Doneda, 2018, p. 112).

Por fim, tendo em vista que segurança de dados é uma ciência “autofágica” (Aranha, 2018), ou seja, seu progresso se dá pela vulnerabilidade de antigas técnicas (a técnica que no determinado momento é posta em prova) e posterior substituição por novas técnicas mais seguras àquele novo cenário, também é importante ao Estado realizar constantes análises de potenciais riscos de identificações em suas bases de dados. **Dados pessoais precisam ser tratados como fato dinâmico, porém não punitivo.** O Estado deve arcar com sua obrigação de tratar e identificar potenciais de risco. No entanto, não pode adotar alternativas simples de restrição às informações ou optar pela disseminando apenas de forma agregada, opções que impossibilitam análises adequadas.

<sup>5</sup> “Art. 5º Para os fins desta Lei, considera-se: I - dado pessoal: informação relacionada a pessoa natural identificada ou identificável”

<sup>6</sup> “Informação pessoal: aquela relacionada à pessoa natural identificada ou identificável”

<sup>7</sup> O Grupo de Proteção de Dados do Artigo 29 foi um grupo independente que trabalhava questões relacionadas à privacidade e proteção de dados até a promulgação do Regulamento Europeu de Proteção de Dados, em maio de 2018, quando é substituído pelo *European Data Protection Board* (EDPB, 2018).



Por fim, se faz importante também entender a hierarquia de acesso aos dados. Uma atividade de monitoramento e avaliação de políticas educacionais trabalhará com dados pseudoanonimizados, por exemplo. Porém, se o que se pretende é o monitoramento de crianças que estão fora da escola é necessário que esses dados sejam acessíveis e a família identificada para determinadas instâncias do poder público. Vale lembrar que é dever do Estado fazer busca ativa das crianças que faltam às aulas, até para identificar causas, como violência doméstica, falta de condições de subsistência, falta de transporte, entre outras possibilidades, de forma a buscar soluções ou minimizar os problemas que levam a essa ausência. Nesse momento, a entidade pública responsável precisa acessar a identidade dessa criança para direcionar sua atuação. Assim, nome, endereço, nome dos pais passam a ser dados necessários para que o Estado exerça a sua obrigação constitucional de salvaguardar as crianças. O direito à privacidade é relativizado em vista do dever de agir estatal.

## Conclusões

O Governo Federal incentiva cada vez mais ações de governo digital, fomentando políticas públicas baseadas em evidência, em que dados estruturados, com séries históricas disponíveis são aspectos chaves para o sucesso das decisões e o processo de implementação da política. Cada vez mais dados possibilitam análises preditivas (através de tendências indicam futuros acontecimentos), prescritivas (sugerem ações para atingir objetivos propostos), descritivas (revisa-se o passado para entender o presente) e diagnósticas (fatores que impactaram acontecimentos passados), como os exemplos descritos ao longo desse texto.

Em um mundo cada vez mais dependente de informação, que estão armazenadas em formato de grandes bancos de dados, é necessário avançar e superar de forma responsável o dilema divulgação de dados e privacidade, visto que não há evidências, transformação digital, cenários futuros, sem consolidação, estruturação e abertura segura de dados. Tendo em mente sempre que informação é poder e em mãos equivocadas podem gerar danos pessoais e coletivos.

A partir desse cenário a pergunta que ascende é: como trabalhar esses dados? Como proteger o usuário sem perder todo o benefício social dos dados armazenados? Qual deve ser o norte desse jogo de preservação da identidade x transparência de dados para políticas públicas? Entendemos que o interesse do cidadão e da sociedade deve ser a régua para avaliar as decisões de acesso a dados. Como regra o Estado deve publicizar e disseminar dados aplicando sofisticadas técnicas de tratamentos de anonimização e pseudononimização às suas bases, além de realizar constantes revisões da segurança das mesmas.

Sendo que o uso da ciência de dados nos mais variados processos deve considerar também a realidade brasileira e suas desigualdades no desenho da política e na necessidade de estabelecer mecanismos cada vez mais transparente de coleta, uso e avaliação de dados sensíveis. Caso contrário, os avanços tecnodigitais vêm para reproduzir iniquidades e, até mesmo, negligências (Filgueiras, 2022). É necessário traduzir o que é feito com os dados pessoais, usando-se também da tecnologia para estabelecer interfaces amigáveis, que facilitem a compreensão do cidadão e/ou da sociedade civil, p.ex. via processos de participação social, considerando os usos feitos e os acessos permitidos.

## Fonte de financiamento

Não há.

## Conflito de interesse

Fernanda Teixeira Reis é editora associada da RBAVAL.



## Referências

- Aranha, Diego F. (2018). O que é criptografia fim a fim e o que devemos fazer a respeito? *Revistas dos Tribunais*, 998, 99-128. (Caderno Especial).
- Banco Mundial. (2021). *Seleção de beneficiários do Auxílio Emergencial: Estratégias de cadastramento e de verificação de elegibilidade*. Washington, DC: World Bank Group. Recuperado em 1 de março de 2023, de <http://documents.worldbank.org/curated/en/099205012022129463/P17483606c29c50ca082ea081e5891caba3>
- Bauer, Raymond. (1967). *Social indicators*. Cambridge: MIT Press.
- Brasil. Ministério da Educação. Instituto Anízio Teixeira. (2021). *Inep publica microdados do Enem 2020 e do Censo Escolar da Educação Básica 2021*. Recuperado em 6 de fevereiro de 2023, de <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/censo-escolar>.
- Brasil. (2022). *Posicionamento Público*. Recuperado em 6 de fevereiro de 2023, de [https://ok.org.br/wp-content/uploads/2022/02/PautasPolíticas\\_Inep\\_Posicionamento\\_MicrodadosEnem\\_2022\\_02\\_22\\_FINAL.pdf](https://ok.org.br/wp-content/uploads/2022/02/PautasPolíticas_Inep_Posicionamento_MicrodadosEnem_2022_02_22_FINAL.pdf)
- Dubois, Vincent. *Contrôler les assistes: Genèses et usages d'un mot d'ordre*. Paris: Raison d'agir; 2021.
- Dulcey, Richard. (2021). *A concise history of the 3G technology*. Recuperado em 6 de fevereiro de 2023, de <https://www.intraway.com/blog/a-concise-history-of-the-3g-technology/>.
- Esayas, Samson Yoseph. (2015). The role of anonymisation and pseudonymisation under the EU data privacy rules: Beyond the 'all or nothing' approach. *The Journal of Law and Technology*, 6(2), 1-23.
- Europa Data Protection Board – EDPB. (2018). *Grupo de Proteção de Dados do Artigo 29*. Recuperado em 6 de fevereiro de 2023, de [https://edpb.europa.eu/about-edpb/who-we-are/legacy-art-29-working-party\\_pt](https://edpb.europa.eu/about-edpb/who-we-are/legacy-art-29-working-party_pt).
- Filgueiras, Fernando. (2022). Big data, artificial intelligence and the future of regulatory tools. In Michael Howlett (Org.), *Routledge handbook on policy tools* (pp. 529-539). London: Routledge.
- Filgueiras, Fernando. (2023). The silent reform: Digital governance as a strategy for state reform in Brazil. In Erika Lisboa, Ricardo Corrêa Gomes & Humberto Falcão Marins (Eds.), *The Brazilian way of doing public administration* (pp. 83-95). Bingley: Emerald.
- Guimarães, José Ribeiro Soares, & Jannuzzi, Paulo de Martino. (2004). Indicadores sintéticos no processo de formulação e avaliação de políticas públicas: limites e legitimidades. In *Encontro Nacional de Estudos Populacionais*, Caxambu (MG). Recuperado em 22 de fevereiro de 2023, de <http://www.abep.org.br/publicacoes/index.php/anais/issue/view/34/showToc>.
- Jornal Extra. (2022). *Robô do INSS indefere 3 de cada 4 pedidos de aposentadoria; veja como aumentar suas chances*. Recuperado em 17 de outubro de 2022, de <https://extra.globo.com/economia-e-financas/robo-doinss-indefere-3-de-cada-4-pedidos-de-aposentadoria-veja-como-aumentar-suas-chances-25588530.html>
- Koga, Natália Massaco, Palotti, Pedro Lucas de Moura, Lins, Rafael da Silva, Couto, Bruno Gontyjo, Loureiro, Miguel, & Lima, Shana Nogueira. (2022). Como os burocratas federais se informam? Uma radiografia das fontes de evidências utilizadas no trabalho das políticas públicas. In Natália Massaco, Koga Pedro Lucas de Moura Palotti, Janine Mello & Maurício Mota Saboya Pinheiro (Eds.), *Políticas públicas e usos de evidências no Brasil: Conceitos, métodos, contextos e práticas*. Brasília, DF: IPEA.
- Machado, Diego, & Doneda, Danilo. (2018). Proteção de dados pessoais e criptografia: tecnologias criptográficas entre anonimização e pseudonimização de dados. *Revistas dos Tribunais*, 998, 99-128.
- Mayer-Schönberger, Vitor, & Cukier, Kenneth. (2013). *Big data: A revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt.
- Paiva, Andrea Barreto, & Pinheiro, Marina Brito. (2021). *BPC em disputa: Como alterações regulatórias recentes se refletem no acesso ao benefício* (Texto para discussão, n. 2703). Brasília, Rio de Janeiro: IPEA. Recuperado em 22 de fevereiro de 2023, de [https://repositorio.ipea.gov.br/bitstream/11058/10890/1/td\\_2703.pdf](https://repositorio.ipea.gov.br/bitstream/11058/10890/1/td_2703.pdf)
- Scandar Neto, Wadih João, Jannuzzi, Paulo Martino, & Silva, Pedro Luís do Nascimento. (2008). Sistemas de indicadores ou indicadores sintéticos: do que precisam os gestores de programas sociais. In *XVI Encontro Nacional de Estudos Populacionais*, Caxambu (MG). Recuperado em 22 de fevereiro de 2023, de <http://www.abep.org.br/publicacoes/index.php/anais/issue/view/36/showToc>.
- Scaratti, Dirceu, & Calvo, Maria Cristina Marino. (2012). Indicador sintético para avaliar a qualidade da gestão municipal da atenção básica à saúde. *Revista de Saúde Pública*, 46(3), 446-455. <http://dx.doi.org/10.1590/S0034-89102012000300006>
- Souza, Celina. (2005). Federalismo, desenho constitucional e instituições federativas no Brasil pós-1988. *Revista de Sociologia e Política*, 24(24), 105-122.