UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

DCC
DEPARTAMENTO DE
CIÊNCIA DA COMPUTAÇÃO

TED 8750 - PRICE

Privacidade nos Censos Educacionais

Termo de Execução Descentralizada entre
o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
e a Universidade Federal de Minas Gerais

Produto 01

Relatório sobre o panorama internacional e o contexto do Inep a respeito dos métodos de tratamento de controle de privacidade na divulgação estatística

INSCRYPT
Laboratory of Information Security,
Cryptography, Privacy, and Transparency

Mário S. Alvim
Jeroen van de Graaf
Ramon G. Gonze
Gabriel H. Nunes

03 de julho de 2020

Resumo

O presente documento consiste em um *Relatório sobre o panorama internacional e o contexto do Inep a respeito dos métodos de tratamento de controle de privacidade na divulgação estatística*, previsto como Produto 01 no plano de trabalho do Termo de Execução Descentralizada (TED) 8750. Este TED foi firmado entre o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) e a Universidade Federal de Minas Gerais (UFMG), e encontra-se em execução pelo Laboratório Inscrypt (*Laboratory of Information Security, Cryptography, Privacy, and Transparency*) do Departamento de Ciência da Computação (DCC) da Universidade.

Neste relatório apresentam-se, através de vários estudos de caso, um panorama internacional e um nacional de como institutos produtores de estatísticas oficiais equilibram requisitos de transparência e de proteção à privacidade em seus métodos de divulgação de informação. Cada estudo de caso consiste em: (i) um sumário da regulamentação geral mais relevante sobre privacidade no país ou região onde o caso se insere; (ii) um sumário da regulamentação específica identificada como pertinente ao instituto em questão; (iii) a identificação dos tipos de dados divulgados pelo instituto, distinguindo-se entre microdados e dados agregados; e (iv) uma descrição das medidas empregadas para mitigação de possíveis violações de privacidade decorrentes da divulgação.

A cobertura do panorama internacional inicia-se com um sumário das principais diretivas e regulamentações gerais sobre privacidade estabelecidas pela Organização das Nações Unidas (ONU) e outros organismos internacionais relevantes. Em seguida, apresentam-se os seguintes estudos de caso no contexto dos Estados Unidos: (1) o caso das práticas de limitação de divulgação das agências federais norte-americanas; (2) o caso da divulgação de dados estatísticos pelo Centro Nacional de Estatísticas da Educação (*National Center for Education Statistics* - NCES); e (3) o caso da divulgação de dados estatísticos pelo Escritório do Censo (*Census Bureau* - USCB); no contexto da União Europeia: (4) o caso da divulgação de dados educacionais por escolas e pelo Ministério da Educação da Holanda; e no contexto da Austrália: (5) o caso da divulgação de dados estatísticos pelo Escritório Australiano de Estatística (*Australian Bureau of Statistics* - ABS).

A cobertura do panorama nacional inicia-se com um sumário das principais diretivas e regulamentações gerais brasileiras sobre transparência e privacidade, com foco na Lei 12.527/2011, conhecida como *Lei de Acesso à Informação* (LAI), e na Lei 13.709/2018,

conhecida como *Lei Geral de Proteção de Dados Pessoais* (LGPDP ou LGPD). Em seguida, apresentam-se seguintes estudos de caso nacionais: (1) o caso da divulgação de dados pelo Instituto Brasileiro de Geografia e Estatística (IBGE); e (2) o caso da divulgação de dados pelo Portal da Transparência do Governo Federal.

Na sequência, procede-se ao cerne deste documento: uma avaliação comparativa da atual forma de divulgação dos Censos Educacionais pelo Inep frente a práticas correntes nos panoramas internacional e nacional, com foco em potenciais riscos à privacidade. Primeiramente, apresenta-se uma contextualização da atual situação do Instituto, incluindo sumários da principal regulamentação específica a ele aplicável, da sua atual forma de divulgação dos Censos Educacionais, e das metodologias de mitigação de danos à privacidade por ele empregadas atualmente. Em seguida, provê-se uma análise comparativa da situação do Inep com aquela dos outros estudos de caso internacionais e nacionais descritos neste documento. A partir desta análise, ressaltam-se as seguintes conclusões:

(i) Frente a todos os casos internacionais estudados, o Inep publica em seus Censos Educacionais a maior quantidade de microdados individuais e em maior nível de detalhes.

(ii) Em todos os casos internacionais estudados, os órgãos produtores de estatísticas oficiais adotam medidas mais rígidas do que as do Inep na mitigação de danos à privacidade causados por publicação de microdados. Tais medidas incluem, e.g., a não publicação de microdados individuais de estudantes sob nenhuma forma, a publicação de microdados apenas por amostras da população, ou a permissão de acesso a todo o conjunto de microdados apenas sob autorização prévia, de acordo com os objetivos do pesquisador ou indivíduo interessado, e em salas seguras.

(iii) As atuais técnicas de proteção de privacidade utilizadas pelo Inep nos microdados divulgados, consistindo apenas em *desidentificação* –em que se removem possíveis identificadores individuais óbvios dos registros, como nome, CPF e RG– e em *pseudonimização* –em que tais identificadores individuais óbvios são substituídos por um código único de identificação artificialmente criado–, estão sujeitas a vários riscos de privacidade já identificados na literatura.

Em seguida, discutem-se riscos decorrentes da atual forma de divulgação adotada pelo Instituto de acordo com a literatura técnica sobre proteção de privacidade.

Por fim, o corpo principal do documento encerra-se com algumas considerações finais.

Como anexos, proveem-se: uma seleção dos trechos da Constituição da República Federativa do Brasil, da Lei de Acesso à Informação (LAI) e da Lei Geral de Proteção de Dados Pessoais (LGPD) identificados como mais relevantes no escopo deste projeto; uma comparação das principais diferenças entre a LGPD brasileira e o Regulamento Geral de Proteção de Dados - RGPD (ou *General Data Protection Regulation* - GDPR) europeu, de acordo com o trabalho de Richie Koch; e um panorama da percepção da questão de privacidade pela sociedade brasileira, formado por uma seleção de reportagens e artigos recentes relevantes produzidos sobre o assunto.

Sumário

Lista de Tabelas

1 Introdução e motivação

Este documento consiste no Produto 01, intitulado *Relatório sobre o panorama internacional e o contexto do Inep a respeito dos métodos de tratamento de controle de privacidade na divulgação estatística*, do plano de trabalho do Termo de Execução Descentralizada (TED) 8750, firmado entre o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) e a Universidade Federal de Minas Gerais (UFMG), e em execução pelo Laboratório Inscrypt (*Laboratory of Information Security, Cryptography, Privacy, and Transparency*) do Departamento de Ciência da Computação (DCC) da Universidade.

1.1 Objetivo deste documento

De acordo com o plano de trabalho do TED 8750, o Produto 01 que este documento implementa consiste em um:

> Relatório com o panorama internacional sobre o tratamento de dados com vistas a garantir a transparência das pesquisas e privacidade dos dados pessoais na divulgação estatística de outros institutos produtores de estatísticas oficiais, incluindo a contextualização da situação do Inep.

Mais especificamente, visa-se a atender ao Objetivo Específico 01 do plano de trabalho:

> [L]evantar o panorama internacional dos procedimentos de garantia da privacidade em bases de dados de pesquisas estatísticas, em especial nos produtos de disseminação dos microdados. O contexto desta pesquisa bibliográfica deve abranger os normativos técnicos e as políticas de tratamento de dados pessoais de outros institutos relevantes de pesquisas estatísticas, no cenário internacional, relacionando-os ao contexto do Inep. O objetivo desta pesquisa é estabelecer a percepção da situação atual do Inep frente a institutos similares em outros países, além de reconhecer os desafios inerentes e as melhores práticas aplicáveis. Neste sentido, a pesquisa será fundamental para a definição das melhorias desejadas na atual política de disseminação de microdados do Inep, uma vez que trará evidências sobre o tratamento técnico realizado por outros institutos e o embasamento necessário à tomada de decisão.

1.2 Contextualização e definição de "privacidade"

Uma primeira aproximação, intuitiva e concisa, da definição do termo "privacidade" pode ser encontrada na Wikipédia [3], que afirma que "privacidade" pode ser entendida como "o direito à reserva de informações pessoais e da própria vida pessoal: o direito de ser deixado em paz". O termo pode ser também entendido como a vontade de controlar a exposição e a disponibilidade de informações acerca de si mesmo, ou seja, a quantidade de controle que um indivíduo exerce sobre a entrada e a saída de declarações de si mesmo e a quantidade de contato que se tem com outras pessoas.

Já uma abordagem mais formal é oferecida pela Dra. Helen Nissenbaum, professora de ciência da informação na Universidade *Cornell Tech* e expoente internacional na área de pesquisa em privacidade. Nissenbaum defende que qualquer interpretação pragmática do termo *privacidade* está atrelada ao fluxo de informações pessoais [78]. Sua tese é de que o que realmente causa a sensação de "violação de privacidade" nas pessoas não é o *ato em si de compartilhamento de informação* –já que a maioria das pessoas entende que isto é essencial para a vida social– mas *o uso inapropriado da informação compartilhada*. Entretanto, os limites e o conteúdo do que é considerado privado diferem entre culturas e indivíduos, e depende também do contexto em que a informação é compartilhada.

A preocupação com a privacidade tem aumentando nas últimas décadas. O Regulamento Geral sobre a Proteção de Dados (RGPD) da União Europeia [29], de 2016, aponta que:

> A rápida evolução tecnológica e a globalização criaram novos desafios em matéria de proteção de dados pessoais. A recolha e a partilha de dados pessoais registaram um aumento significativo. As novas tecnologias permitem às empresas privadas e às entidades públicas a utilização de dados pessoais numa escala sem precedentes no exercício das suas atividades. As pessoas singulares disponibilizam cada vez mais as suas informações pessoais de uma forma pública e global. As novas tecnologias transformaram a economia e a vida social e deverão contribuir para facilitar a livre circulação de dados pessoais na União e a sua transferência para países terceiros e organizações internacionais, assegurando simultaneamente um elevado nível de proteção dos dados pessoais.

No contexto tecnológico, tal preocupação já existe há pelo menos algumas dezenas de anos. Por exemplo, já no ano 2000 Simson e Garfinkel descreviam em seu livro *Database Nation* [56] diversos cenários de possíveis violações de privacidade decorrentes do uso de novas tecnologias disponíveis na época. Os exemplos citados incluíam casos em que dados de um consumidor poderiam ser cruzados (e.g., dados do cartão de crédito com a conta do supermercado), ou em que *cookies* em navegadores poderiam ser usados para rastreamento do comportamento do usuário na Internet.

Como um exemplo de como o uso de novas tecnologias pode causar a sensação de violação de privacidade de maneiras inéditas –e que desafiam nossa capacidade de classificar precisamente onde teria ocorrido a efetiva "violação"–, considere o seguinte cenário. Imagine o

centro de uma cidade onde em toda esquina haja câmeras que gravem a imagem dos transeuntes e os identifiquem por reconhecimento facial ou pelo seu "modo de andar" (*gait*, no termo original em inglês). Considere ainda que todas as câmeras estejam conectadas a um grande banco de dados, possibilitando o cruzamento de informações e permitindo, assim, o acompanhamento em tempo real dos trajetos realizados por qualquer indivíduo pela cidade. A contemplação de tal cenário provoca em um grande número de pessoas um desconforto associado a uma percepção de *violação de privacidade*. Repare, entretanto, que *toda a informação coletada neste caso consiste, em princípio, em informação pública*, afinal, o trajeto de pessoas em vias públicas é observável por qualquer um que esteja nestas vias. Porém, sem aparatos tecnológicos, seria impraticável para agentes do Estado parar cada pessoa em cada esquina e registrar seus movimentos, e seria ainda mais infactível processar e cruzar as informações obtidas em tempo real.

O exemplo anterior ilustra um ponto crucial da realidade em que vivemos, o qual é central para o presente projeto: *o uso de tecnologia possibilita que informações a princípio consideradas públicas se tornem uma ameaça à privacidade de indivíduos*. Consequentemente, a crescente capacidade de coleta, armazenamento e processamento de informação provida por avanços tecnológicos força a sociedade a continuamente rever sua percepção de riscos de violação de privacidade. Do ponto de vista prático, no ano de 2020 o cenário esboçado no exemplo acima nem pode mais ser considerado puramente futurista: a cidade de Londres, na Inglaterra, já analisa as imagens de milhares de câmeras para proteger seu centro contra ataques terroristas [49], enquanto o governo da China usa essa mesma tecnologia para monitorar sua população [55]. Em ambos os casos, há forte reação de setores da sociedade preocupados com a proteção da privacidade dos cidadãos.

No Brasil, a preocupação com privacidade permaneceu relativamente diminuta até recentemente. De fato, frequentemente vinham-se adotando novas tecnologias sem um cuidadoso senso crítico quanto aos potenciais riscos à privacidade delas decorrentes. Esse cenário, entretanto, vem felizmente evoluindo. Pode-se argumentar que um catalisador relevante para essa maior conscientização sobre os riscos das novas tecnologias digitais em relação à privacidade foram as revelações de Edward Snowden em 2013. Snowden demonstrou que o governo dos EUA colecionava dados de cidadãos estrangeiros numa escala gigantesca e sem precedentes, desenvolvendo plataformas e tecnologias específicas para cruzar e minerar esses dados [16]. No caso particular do Brasil, chamou a atenção a revelação de que os celulares da então presidente da República, Dilma Rousseff, e de seus assessores foram *hackeados*, vazando dados dos círculos mais restritos do governo.

A conscientização fomentada por estes e outros acontecimentos eventualmente provocou respostas do Poder Legislativo brasileiro. A Lei 13.709/2018, conhecida como *Lei Geral de Proteção de Dados Pessoais* (LGPDP ou LGPD) [41], sancionada em 14 de agosto de 2018 e publicada no Diário Oficial da União no dia seguinte, alterou significativamente a maneira como organismos públicos e privados devem lidar com informações pessoais, visando à preservação da privacidade. [1]

[1] A princípio, o início da vigência da LGPD ocorreria em fevereiro de 2020. No entanto, em dezembro de 2018, o presidente Michel Temer editou a Medida Provisória 869/2018 [61], prevendo a criação da

Por outro lado, a Lei 12.527/2011, conhecida como *Lei de Acesso à Informação* (LAI) [47], determina ao poder público o dever de garantir o amplo acesso à informação, particularmente àquela considerada de interesse coletivo ou geral, a qual deve ser disponibilizada via Internet independentemente de requerimento. Determina, também, o direito de acesso às demais informações não imediatamente disponibilizadas via Internet, e garante o acesso à informação primária, íntegra, autêntica, e atualizada em ambos os casos. [2] A própria LAI aponta, entretanto, que a informação pessoal deve ser protegida, reconhecendo que o cuidado com a privacidade dos indivíduos é uma questão central mesmo na regulamentação sobre acesso a dados.

Neste contexto, impõe-se aos órgãos públicos que coletam, mantêm e divulgam dados o desafio de conciliar os requisitos de transparência com os requisitos de proteção de privacidade impostos pela atual legislação brasileira. Na próxima seção, introduzimos tal desafio no contexto particular do Inep relativo à divulgação de seus Censos Educacionais.

1.3 A preocupação com privacidade relativa à divulgação pelo Inep dos Censos Educacionais

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) [3], autarquia federal vinculada ao Ministério da Educação (MEC), [4] tem entre suas atribuições a responsabilidade pela execução de levantamentos estatísticos sobre a educação nacional nos níveis básico e superior, incluindo o tratamento de dados, o controle de qualidade, e a produção e divulgação das estatísticas anuais específicas e dos indicadores educacionais.

Nesta seção, fundamentamos a relevância do presente TED 8750, que tem como objetivo uma cuidadosa avaliação da atual forma como o Inep lida com o compromisso entre transparência na divulgação de dados de interesse público e a proteção à privacidade dos titulares dos dados. O contexto do Inep é discutido em detalhes na Seção 4, mas para os objetivos da presente seção é suficiente notar os seguintes pontos:

1. O Inep produz e divulga várias Pesquisas de grande relevância nacional, incluindo o *Censo da Educação Básica*, o *Censo da Educação Superior*, o *Exame Nacional do Ensino Médio* (ENEM), o *Sistema de Avaliação da Educação Básica* (SAEB), o *Exame Nacional de Certificação de Competências de Jovens e Adultos* (ENCCEJA) e o *Exame Nacional de Desempenho dos Estudantes* (ENADE).

Autoridade Nacional de Proteção de Dados (ANPD) para fiscalizar o cumprimento da LGPD e alterando o início da vigência da lei para 14 de agosto de 2020. Em abril de 2020, o presidente Jair Bolsonaro editou a Medida Provisória 959/2020 [63], adiando novamente a entrada em vigor da LGPD para maio de 2021. Finalmente, em junho de 2020 foi sancionada a Lei 14.010/2020, que determina que as sanções da LPGD só devem ser aplicadas a partir de agosto de 2021 [42]. A LGPD é abordada em mais detalhes na Seção 3.1.

[2] A LAI também é abordada em mais detalhes na Seção 3.1.

[3] http://www.inep.gov.br/

[4] https://www.gov.br/mec/

2. O formato atual de publicação destes Censos consiste em *microdados*, ou seja, dados na menor unidade de agregação possível, que incluem informações em nível individual sobre alunos, professores e escolas.

3. As técnicas de proteção de privacidade atualmente aplicadas aos microdados divulgados pelo Inep consistem em:

    - *desidentificação*, pela qual são removidos possíveis identificadores individuais óbvios dos registros (como nome, CPF, RG, ou endereços em níveis mais detalhados que as cidades); e
    - *pseudonimização*, em que o Instituto atribui a cada registro um código único de identificação artificialmente criado, substituindo identificadores individuais naturais (como nome, CPF ou RG).

Neste contexto, a harmonização dos requisitos legais da Lei de Acesso à Informação (LAI) e da Lei Geral de Proteção de Dados Pessoais (LGPD) se apresentam como desafios para o Inep, em particular, na divulgação dos dados de suas Pesquisas.

De fato, os pesquisadores brasileiros Queiroz e Motta já apontaram, em estudo de 2015, fragilidades quanto à proteção de privacidade decorrentes da atual forma de divulgação de dados dos Censos Educacionais pelo Inep [102]. Utilizando o arquivo referente aos docentes do Censo da Educação Superior do ano de 2013, os pesquisadores puderam reidentificar unicamente um dos autores do estudo dentre todos os 383.683 registros de docentes presentes na base, utilizando apenas sua data de nascimento, gênero, e nome da Instituição de Ensino Superior à qual estava vinculado.

O Produto 02 do presente TED 8750 tem como objetivo principal produzir uma análise quantitativa deste e de outros riscos de violação de privacidade no contexto do Inep. Além disso, a Seção 4.4 do presente documento apresenta uma análise preliminar dos potenciais riscos à privacidade previstos pela literatura técnica na área, além de uma contextualização da atual situação do Inep frente a outros institutos de estatísticas oficiais inseridos nos panoramas internacional e nacional. Na presente seção, portanto, limitamo-nos a produzir mais um exemplo de risco à privacidade que julgamos ilustrativo.

O filho de um dos autores do presente documento tem hoje 14 anos e, portanto, consta nas divulgações dos microdados dos Censos do Ensino Básico pelo Inep em vários anos distintos. Apesar do nome, RG e CPF do adolescente não constarem nos microdados publicados, pudemos reidentificá-lo de forma relativamente fácil da seguinte maneira. Primeiramente, a partir do nome da escola onde o adolescente estuda, encontramos o identificador único correspondente a esta escola na própria base do Inep. Em um segundo momento, entre os alunos desta escola selecionamos na base aqueles com a mesma data de nascimento do adolescente em quem estávamos interessados. Como havia apenas um estudante nesta escola com a data de nascimento em questão, e sabendo que o adolescente alvo estava na base, pudemos reidentificá-lo e recuperar seus dados sensíveis, como, por exemplo, se ele é ou não portador de necessidades especiais.

Neste caso, a reidentificação foi possível porque não havia nenhum outro aluno naquela mesma escola com a mesma data de nascimento do aluno alvo. Caso houvesse, a inspeção dos microdados retornaria o registro de todos estes alunos, e seria necessária alguma informação auxiliar (como a série do aluno) para reidentificá-lo. Porém, nesse ponto, seria possível que alguma informação sensível sobre o adolescente já tivesse sido revelada. Por exemplo, caso todos os alunos que compartilhem de sua data de nascimento em sua escola fossem (ou não fossem) portadores de necessidades especiais, ficaria evidenciado se o aluno em questão também seria (ou não seria) portador de tais necessidades. Este exemplo demonstra que *mesmo que o aluno não tivesse sido unicamente reidentificado na base do Inep, uma violação de privacidade já poderia ter ocorrido.*

No exemplo acima, a reidentificação desse aluno em particular demandou menos de 20 minutos de trabalho de um membro da equipe do Inscrypt que já está familiarizado com as bases de microdados do Censo da Educação Básica. Entretanto, seria perfeitamente possível implementar-se uma página na Internet onde o usuário escolheria o nome da escola, digitaria a data de nascimento de um aluno, e visualizaria imediatamente todos os dados do aluno presentes no Censo. O esforço para desenvolver tal página não seria proibitivo: talvez um dia de trabalho de um profissional de tecnologia da informação.

A discussão acima demonstra que há uma possibilidade concreta de violações de privacidade decorrentes da atual forma de divulgação dos Censos Educacionais por parte do Inep. É imperativo, portanto, estudar cuidadosamente o contexto de tais Censos, identificando suas peculiaridades e encontrando técnicas que permitam o melhor compromisso entre a transparência da informação divulgada e a proteção da privacidade dos indivíduos. Este é exatamente o objetivo geral do presente TED 8750.

1.4 Organização deste documento

O presente documento se foca em estudos de caso de como outras instituições produtoras de estatísticas, em nível internacional e nacional, atuam na mitigação de possíveis danos à privacidade, e na comparação destes estudos de caso com a atual situação do Inep.

O restante deste documento está organizado da seguinte forma.

- **Estudos de caso do panorama internacional** (Seção 2): Nesta seção provemos um panorama internacional de como institutos produtores de estatísticas oficiais equilibram requisitos de transparência e de proteção à privacidade em seus métodos de divulgação de informação. Iniciamos a seção com um sumário das principais diretivas e regulamentações gerais sobre privacidade estabelecidas pela Organização das Nações Unidas (ONU) e outros organismos internacionais relevantes (Seção 2.1). Em seguida, apresentamos a seguinte coleção de estudos de caso internacionais.

    I. **No contexto dos Estados Unidos da América** (Seção 2.2): Após um sumário dos aspectos mais relevantes da regulamentação geral norte-americana sobre privacidade (Seção 2.2.1), apresentamos:

  - **Estudo de caso internacional 1** (Seção 2.2.2): *Práticas de limitação de divulgação das agências federais norte-americanas.*
  - **Estudo de caso internacional 2** (Seção 2.2.3): *Divulgação de dados estatísticos pelo Centro Nacional de Estatísticas da Educação* (National Center for Education Statistics - NCES).
  - **Estudo de caso internacional 3** (Seção 2.2.4): *Divulgação de dados estatísticos pelo Escritório do Censo* (Census Bureau - USCB).

II. **No contexto da União Europeia** (Seção 2.3): Após um sumário dos aspectos mais relevantes da regulamentação geral da União Europeia sobre privacidade (Seção 2.3.1), apresentamos:
  - **Estudo de caso internacional 4** (Seção 2.3.2): *Divulgação de dados educacionais por escolas e pelo Ministério da Educação da Holanda.*

III. **No contexto da Austrália** (Seção 2.4): Após um sumário dos aspectos mais relevantes da regulamentação geral australiana sobre privacidade (Seção 2.4.1), apresentamos:
  - **Estudo de caso internacional 5** (Seção 2.4.2): *Divulgação de dados estatísticos pelo Escritório Australiano de Estatística* (Australian Bureau of Statistics - ABS).

- **Estudos de caso do panorama nacional** (Seção 3): Nesta seção provemos um panorama nacional de como institutos produtores de estatísticas oficiais equilibram requisitos de transparência e de proteção de privacidade em seus métodos de divulgação de informação. Iniciamos a seção com um sumário das principais diretivas e regulamentações gerais brasileiras sobre transparência e privacidade (Seção 3.1), com especial foco na Lei de Acesso à Informação (LAI) e na Lei Geral de Proteção de Dados Pessoais (LGPD), além do Marco Civil da Internet (MCI). Em seguida, apresentamos a seguinte coleção de estudos de caso nacionais.
  - **Estudo de caso nacional 1** (Seção 3.2): *Divulgação de dados pelo Instituto Brasileiro de Geografia e Estatística (IBGE).*
  - **Estudo de caso nacional 2** (Seção 3.3): *Divulgação de dados pelo Portal da Transparência do Governo Federal.*

- **Análise da atual situação do Inep frente aos panoramas internacional e nacional e à literatura técnica em privacidade** (Seção 4): Nesta seção apresentamos uma análise da atual forma de divulgação dos Censos Educacionais por parte do Inep, com especial foco nos possíveis riscos à privacidade dela decorrentes. Iniciamos a seção com uma contextualização da atual situação do Instituto (Seção 4.1), que inclui um sumário dos principais pontos da regulamentação específica aplicável ao mesmo (Seção 4.1.1), uma breve descrição da atual forma de divulgação dos Censos Educacionais realizados pelo Inep (Seção 4.1.2) e a identificação das atuais metodologias de mitigação de danos à privacidade aplicadas pelo Instituto (Seção 4.1.3). Em seguida, apresentamos um sumário comparativo do contexto do

Inep com o dos outros casos de estudo internacionais e nacionais já apresentados neste documento (Seção 4.2) e destacamos os pontos identificados como os mais relevantes à atual situação do Inep (Seção 4.3). Por fim, realizamos uma avaliação preliminar dos possíveis riscos decorrentes da forma de divulgação adotada pelo Instituto de acordo com o conhecimento acumulado na literatura técnica sobre proteção de privacidade (Seção 4.4).

- **Considerações finais** (Seção 5): Nesta seção apresentamos as considerações finais deste documento.

- **Anexos**: Apresentamos aqui o seguinte material suplementar.

  - **Constituição da República Federativa do Brasil: Principais trechos relacionados ao escopo deste projeto** (Anexo A): Neste anexo apresentamos uma seleção dos trechos da Constituição da República Federativa do Brasil identificados como os mais relevantes no escopo do presente projeto.
  - **Lei 12.527/2011 (LAI): Principais trechos relacionados ao escopo deste projeto** (Anexo B): Neste anexo apresentamos uma seleção dos trechos da Lei de Acesso à Informação (LAI) identificados como os mais relevantes no escopo do presente projeto.
  - **Lei 13.709/2018 (LGPD): Principais trechos relacionados ao escopo deste projeto** (Anexo C): Neste anexo apresentamos uma seleção dos trechos da Lei Geral de Proteção de Dados Pessoais (LGPD) identificados como os mais relevantes no escopo do presente projeto.
  - **Principais diferenças entre a LGPD brasileira e a GDPR europeia** (Anexo D): Neste anexo apresentamos uma seleção das principais diferenças entre a Lei Geral de Proteção de Dados Pessoais (LGPD) brasileira e o Regulamento Geral de Proteção de Dados - RGPD (*General Data Protection Regulation* - GDPR) europeu, de acordo com o trabalho de Richie Koch.
  - **Um panorama da percepção da questão de privacidade pela sociedade brasileira** (Anexo E): Neste anexo apresentamos uma coleção de notícias e artigos relevantes recentes que ilustram a percepção da sociedade brasileira quanto a questões de privacidade.

- **Lista de siglas e Glossário** encerram o documento. Recomendamos ao leitor a consulta ao Glossário para encontrar a definição de termos utilizados recorrentemente neste documento, como, por exemplo, *anonimização*, *pseudonimização*, *desidentificação*, *reidentificação*, dentre outros.

2 Estudos de caso do panorama internacional

Nesta seção provemos um panorama internacional dos diferentes métodos pelos quais institutos produtores de estatísticas oficiais divulgam informação de forma a garantir transparência e, ao mesmo tempo, privacidade de dados pessoais. Iniciamos a seção com um sumário das principais diretivas e regulamentações gerais sobre privacidade estabelecidas pela Organização das Nações Unidas (ONU) e outros organismos internacionais relevantes. Em seguida, apresentamos uma coleção de estudos de caso no contexto dos Estados Unidos, União Europeia e Austrália.

2.1 Contexto da Organização das Nações Unidas e de outros organismos internacionais

Em 29 de janeiro de 2014, a Assembleia Geral da Organização das Nações Unidas (ONU) adotou a Resolução 68/261, que estabelece um conjunto de princípios fundamentais para estatísticas oficiais [30]. Os princípios definidos nessa Resolução advêm de um trabalho desenvolvido pela Conferência dos Estatísticos Europeus em 1992 e adotado pela primeira vez pela Comissão de Estatística das Nações Unidas em abril de 1994 [32].

Dentre as motivações da Assembleia Geral da ONU para a adoção desses princípios está a necessidade de que o público confie nos sistemas estatísticos oficiais, o que, de acordo com o preâmbulo da Resolução, decorre do respeito aos valores e princípios fundamentais que nela constam. Por sua vez, dentre os princípios estabelecidos, destaca-se:

> Princípio 6. Os dados individuais coletados pelos órgãos estatísticos para a elaboração de estatísticas, sejam eles referentes a pessoas físicas ou jurídicas, devem ser estritamente confidenciais e utilizados exclusivamente para fins estatísticos.

Anteriormente à adoção dos princípios pela Assembleia Geral da ONU, ainda quando eles haviam sido adotados apenas pela Comissão de Estatística da entidade, os mesmos

serviram de base para que o Serviço de Estatística da União Europeia (EUROSTAT) elaborasse a *Código de Conduta para as Estatísticas Europeias* em 2001, revisado em 2017 [53]. Posteriormente, a declaração foi utilizada para o desenvolvimento do *Código de Boas Práticas das Estatísticas Europeias*, adaptado à realidade regional, adotado em 2005, e revisado em 2019 [54, 33].

Similarmente, a Conferência Estatística das Américas (CEA) da Comissão Econômica para a América Latina e o Caribe (CEPAL) desenvolveu, entre os anos de 2009 e 2011, o *Código Regional de Boas Práticas das Estatísticas para a América Latina e o Caribe*, aprovado em Novembro de 2011 durante a Sexta Reunião da CEA [27]. O desenvolvimento do Código Regional contou com a colaboração do EUROSTAT e de 14 países membros da CEA-CEPAL, dentre eles Argentina, Brasil, Paraguai, Uruguai, e Chile. [1] Entretanto, essas Resoluções, Declarações, ou Códigos apenas estabelecem princípios e boas práticas nos quais órgãos estatísticos governamentais deveriam se basear, sem que, na prática, tenham força de Lei.

Já no contexto comercial, é sabido que o Brasil aspira a associar-se à Organização para a Cooperação e o Desenvolvimento Econômico (OCDE). Segundo o Itamaraty [75]:

> A Organização para a Cooperação e o Desenvolvimento Econômico (OCDE) constitui foro composto por 35 países, dedicado à promoção de padrões convergentes em vários temas, como questões econômicas, financeiras, comerciais, sociais e ambientais. Suas reuniões e debates permitem troca de experiências e coordenação de políticas em áreas diversas da atuação governamental.
>
> [...]
>
> Em junho de 2015, o Brasil e a OCDE assinaram um acordo de cooperação, que permitirá aprofundar e sistematizar o relacionamento bilateral. O acordo institucionaliza a participação brasileira em diversos foros da OCDE e estabelece mecanismos para a definição de linhas de trabalho futuras.

Para possibilitar o fluxo livre de informações entre países, considerado essencial na nova economia digital, a OCDE zela para que a legislação relativa à privacidade seja o mais uniforme possível em seus países membros. Para tanto, a Organização criou, em 2013, um documento descrevendo seu arcabouço de privacidade [96]. Esse documento serviu como base para a legislação brasileira que veio posteriormente. Além disso, a OCDE mantém vários documentos tratando especificamente dos riscos enfrentados por crianças no ambiente *on-line* [97, 104].

[1] Posteriormente, esse Código Regional serviu como base para que o Instituto Brasileiro de Geografia e Estatística (IBGE) elaborasse o *Código de Boas Práticas das Estatísticas do IBGE*, publicado em 2013 [33], e discutido na Seção 3.2. Por sua vez, tanto o Código Regional quanto o do IBGE serviram de orientação para a Portaria 91/2017 do Inep [64], que versa sobre o mesmo tema no contexto do Instituto, e para a Portaria 492/2018 [65], que institui a *Política de Divulgação de Estatísticas, Exames e Avaliações, Estudos e Pesquisas do Inep.*. O contexto do Inep é discutido em mais detalhes na Seção 4.

2.2 Contexto dos Estados Unidos da América

Nesta seção, após um sumário dos aspectos mais relevantes da regulamentação geral norte-americana sobre privacidade, apresentamos três estudos de caso no contexto dos Estados Unidos da América (EUA).

2.2.1 Regulamentação geral

Os Estados Unidos da América criaram sua primeira *Lei de Privacidade* em 1974, em uma emenda ao Código dos Estados Unidos que acrescentou a Seção 552a ao seu Título 5 [89]. De acordo com a nova seção, agências federais que coletam, mantêm, usam, ou disseminam qualquer registro de informações pessoais identificáveis devem certificar-se de que essas ações ocorram de forma legal, e que sejam fornecidas salvaguardas adequadas para evitar seu uso indevido. Particularmente no uso desses registros para pesquisas estatísticas, a transferência das informações deve dar-se de tal forma a garantir que os registros não sejam individualmente identificáveis, conforme a seção 552a(b)(5) [95].

Em 2002, a *Lei de Proteção à Informação Confidencial e Eficiência Estatística* [92] foi aprovada e trouxe em seu texto novas salvaguardas às informações individualmente identificáveis fornecidas a agências federais para fins estatísticos sob promessa de confidencialidade. A partir de então, qualquer divulgação intencional de tais informações para fins não estatísticos, sem o consentimento do titular, tornou-se crime federal.

Diversas agências federais dos Estados Unidos coletam e divulgam informações pessoais para fins estatísticos, dentre elas o Centro Nacional de Estatísticas da Educação (*National Center for Education Statistics* - NCES) e o Escritório do Censo (*United States Census Bureau* - USCB). Vale destacar que cada agência está subordinada a um Departamento específico do governo federal e sujeita às normas estabelecidas por ele, além de que, em alguns casos, há também legislação específica que rege o mandato de certas agências.

2.2.2 Estudo de caso internacional 1: Práticas de limitação de divulgação das agências federais

Originalmente publicado em 1994, o *Documento de Trabalho sobre Política Estatística 22 - Relatório sobre a Metodologia de Limitação de Divulgação Estatística* foi revisado em 2005 [93] e traz não apenas definições de diversos métodos de limitação de divulgação estatística, como também as práticas adotadas à época pelas seguintes 14 agências federais dos Estados Unidos.

- *Economic Research Service* (ERS), responsável por divulgar dados sobre agricultura e economia.
- *National Agricultural Statistics Service* (NASS), responsável por divulgar dados sobre produção agrícola, economia, demografia, e meio ambiente.

- *Bureau of Economic Analysis* (BEA), responsável por divulgar dados sobre macroeconomia e indústria, incluindo o produto interno bruto nacional e regional.
- *United States Census Bureau* (USCB), responsável por divulgar dados sobre população e economia.
- *National Center for Education Statistics* (NCES), responsável por divulgar dados sobre educação e informações sobre finanças de distritos escolares.
- *Energy Information Administration* (EIA), responsável por divulgar dados sobre energia, incluindo dados sobre carvão, petróleo, gás natural, energia elétrica, renovável, e energia nuclear.
- *National Center for Health Statistics* (NCHS), responsável por divulgar dados sobre saúde pública.
- *Agency for Healthcare Research & Quality* (AHRQ), responsável por divulgar dados sobre qualidade, adequação e eficácia dos serviços de saúde, assim como o acesso aos mesmos.
- *Social Security Administration* (SSA), responsável por divulgar dados sobre renda e uma lista anual com os nomes mais comumente dados a bebês recém-nascidos nos Estados Unidos no ano anterior.
- *Bureau of Justice Statistics* (BJS), responsável por divulgar dados sobre criminalidade.
- *Bureau of Labor Statistics* (BLS), responsável por divulgar dados sobre economia e trabalho.
- *Internal Revenue Service, Statistics of Income Division* (IRS), responsável por divulgar dados sobre declarações de imposto de renda e informações relacionadas.
- *Bureau of Transportation Statistics* (BTS), responsável por divulgar dados sobre transporte.
- *National Science Foundation* (NSF), responsável por divulgar dados sobre educação e pesquisa básica em todos os campos não médicos da ciência e engenharia.

O Documento 22 descreve que as agências federais acima adotavam em 2005 como formas de divulgação de informação:

- *Microdados*: tipo de dados em nível de registros individuais.
- *Dados agregados*, incluindo:
  - *Dados de frequência*: tipo de dados relacionado a contagens, e.g., quantos estabelecimentos estão operando em um determinado estado.

– *Dados de magnitude*: tipo de dados relacionado a grandezas como lucro, e.g., contagem de estabelecimentos juntamente com a receita bruta agregada.

No que se refere à publicação de microdados, o Documento 22 reconhece a dificuldade inerente de proteger-se esse tipo de registro contra violações de privacidade, uma vez que há sempre a possibilidade de que informações externas aos microdados sejam utilizadas para a reidentificação de indivíduos. Ainda assim, algumas agências dos Estados Unidos publicavam microdados à época da elaboração do Documento 22, sendo que todas elas utilizavam os seguintes métodos de proteção:

(i) inclusão de registros de apenas uma amostra da população;

(ii) não inclusão de identificadores óbvios, como nome, endereço e números de identificação;

(iii) limitação do detalhamento geográfico, i.e., agregar registros em regiões mais amplas; e

(iv) limitação da quantidade e do detalhamento de categorias dentro das variáveis.

Particularmente para variáveis consideradas de elevado risco, como renda, etnia e idade, os seguintes métodos também eram utilizados:

(v) *top-* e *bottom-coding*, i.e., truncamento de códigos extremos para certas variáveis, não revelando valores exatos acima ou abaixo de um limite;

(vi) recodificação em intervalos ou arredondamentos;

(vii) adição ou multiplicação por números aleatórios (ruído), e.g., a partir de uma distribuição normal;

(viii) permutação, i.e., selecionar uma amostra, encontrar uma correspondência dentre os registros restantes dado um conjunto de variáveis, e permutar as demais variáveis entre os grupos;

(ix) troca de classificação, o que permite a seleção de pares para permutação em variáveis contínuas, uma vez que os registros estejam próximos na classificação;

(x) apagamento e imputação, i.e., seleção de registros aleatoriamente, apagando variáveis selecionadas e atribuindo novos valores a elas com base em um modelo;

(xi) *blurring*, i.e., agregamento em pequenos grupos, substituindo o valor relatado de um registro pela média; e

(xii) supressão direcionada, i.e., remoção de um registro por completo ou de valores sensíveis para certas variáveis no registro caso não seja possível protegê-lo de outra forma.

Vale ressaltar que os microdados submetidos à permutação, ao método de apagamento e imputação, ou ao *blurring* não são marcados de qualquer forma nos arquivos divulgados, de modo a garantir a segurança desses registros. Entretanto, isso adiciona um elevado grau de incerteza aos dados, uma vez que o usuário se torna incapaz de determinar se um registro isolado contém dados reais, permutados, imputados, ou ofuscados.

Além disso, tanto a divulgação de microdados quanto a atualização de dados divulgados e a inclusão de dados adicionais devem ser previamente determinadas pelos seguintes órgãos:

- *Disclosure Review Board (DRB)*: este Conselho de Revisão de Divulgação funciona como um curador que determina a liberação ou não de novos conjuntos de microdados, tendo em vista fatores como a disponibilidade de arquivos externos com dados comparáveis, esforço computacional estimado para reidentificação, sensibilidade de itens individuais, quantidade de registros únicos no conjunto de dados, proporção da população incluída na amostra, erros esperados nos dados, e idade do conjunto de dados.

- *CENSUS*: o Escritório do Censo (USCB) é autorizado a conduzir pesquisas para outras agências governamentais, conforme os Títulos 13 e 15 do Código dos Estados Unidos [88, 90]. Quaisquer dados colhidos dessa forma estão sob a jurisdição especial dada ao USCB, e os procedimentos de limitação de divulgação estatística utilizados são então determinados pelo Escritório do Censo, e não pela agência responsável pela publicação.

Já no que se refere à publicação de dados agregados, o Documento 22 reconhece que a simples agregação não garante necessariamente a proteção dos mesmos, e diversas técnicas aplicáveis a microdados já eram aplicadas à época também a dados agregados de magnitude e de frequência.

Por fim, o Escritório do Censo (USCB) disponibilizou também um sistema de consultas (ou *queries*) avançado (*Advanced Query System*) *on-line* para acesso aos dados do Censo do ano 2000. Nesse sistema, o usuário podia definir consultas próprias que então eram analisadas por um primeiro filtro (*Query Filter*), o qual identificava se a consulta enviada revelaria ou não mais informações do que o permitido, e.g., por extrapolar limites geográficos. Caso a consulta fosse aprovada pelo primeiro filtro, a tabela gerada por ela era então submetida a um segundo filtro, agora de resultados estatísticos (*Statistical Results Filter*), responsável por verificar condições de limitação de divulgação de dados nas células da tabela. Finalmente, a tabela resultante era fornecida ao usuário apenas caso as condições fossem satisfeitas.

A Tabela 2.2.1 apresenta um resumo das práticas de proteção de privacidade adotadas pelas agências federais dos Estados Unidos em 2005 [93], a todos os tipos de dados (microdados e dados agregados de frequência e magnitude), onde os seguintes termos adicionais são usados:

Agência	Tipo de Dados			Autorização Individual	Acesso Restrito
	Magnitude	Frequência	Microdados		
ERS	❶ ❸	❺ ❸	✗	✓	✓
NASS	❶ ❷	✗	✗	✓	✓
BEA	❷	✗	✗	✗	✓
USCB	❷ ❽	❺ ❻ ❾	✓ (DRB)	✓	✓
NCES	❸ ❺ ❻ ❼	❸ ❺ ❻ ❼	✓ (DRB)	✗	✓
EIA	❶	❺	✓ (DRB)	✓	✗
NCHS	❶	❹ ❺	✓ (DRB)	✗	✓
AHRQ	✗	❹ ❺	✓ (DRB)	✓	✓
SSA	❸ ❺	❸ ❺	✓ (DRB)	✗	✗
BJS	✗	❺ ❿	✓ (DRB)	✗	✗
BLS	❶ ❷	variável	✓ (CENSUS)	✓	✓
IRS	❸ ❺	❸ ❺	✓ (DRB)	✗	✗
BTS	variável	❸ ❺	✓ (DRB)	✗	✗
NSF	❶	variável	✓ (DRB)	✓	✓

Tabela 2.2.1: Práticas de proteção à privacidade adotadas pelas agências federais norte-americanas em 2005: ❶ (n,k)-*rule*, ❷ *p-percent rule*, ❸ 3+, ❹ 4+, ❺ truncamento, ❻ permutação, ❼ recodificação, ❽ ruído, ❾ sistema avançado de consultas, ❿ 10+.

- Níveis de acesso a dados possíveis:
  - *Autorização individual*: permissão por escrito obtida diretamente dos entrevistados que autoriza a publicação de células sensíveis como uma forma de evitar a supressão de dados.
  - *Acesso restrito*: sala de análise segura pertencente à agência federal para pesquisadores visitantes ou licenciamento externo.
- Critérios de determinação de se uma célula é sensível:
  - (n,k)-*rule*: independentemente da quantidade de registros em uma célula da tabela, se uma pequena quantidade (n ou menor que n) contribui em uma grande porcentagem (k porcento ou mais) para o valor total da célula, então a célula é considerada sensível.
  - *p-percent rule*: uma célula é considerada sensível se for possível estimar o valor de algum dos indivíduos que a compõem dentro de uma margem determinada por um valor p em torno do valor real.
  - *Threshold rule (N+)*: uma célula é considerada sensível se a quantidade de indivíduos reportados por ela for inferior a um tamanho mínimo N determinado (e.g., para 3+, 4+, ou 10+, cada célula representa ao menos 3, 4, ou 10 indivíduos, respectivamente).

2.2.3 Estudo de caso internacional 2: Divulgação de dados estatísticos pelo Centro Nacional de Estatísticas da Educação

Na seção anterior, cobrimos o panorama geral das agências federais dos EUA; aqui dedicamos especial atenção a uma agência com particular relevância no contexto do presente projeto: O Centro Nacional de Estatísticas da Educação (*National Center for Education Statistics* - NCES). [2] Como uma agência estatística do Departamento de Educação dos Estados Unidos (*United States Department of Education* - ED), [3] o NCES é o principal responsável pela coleta, análise e divulgação de dados educacionais no país. Dentre os dados coletados, estão aqueles referentes às instituições de ensino e respectivos corpos docentes, administrativos, e discentes, sendo as publicações realizadas por meio de dados agregados, como em tabelas de frequência ou magnitude, e por meio de microdados.

Regulamentação específica Assim como as demais agências federais, o NCES é regido pela *Lei de Privacidade* de 1974 e a *Lei de Proteção à Informação Confidencial e Eficiência Estatística* de 2002. Entretanto, certa regulamentação específica também se aplica ao mandato do NCES.

A *Lei de Reforma das Ciências da Educação* de 2002 [87], dentre outras providências, trata especificamente sobre os trabalhos do NCES. A respeito das informações individualmente identificáveis sobre estudantes, suas famílias, e escolas, a Lei determina que todas devem permanecer confidenciais.

Entretanto, o *USA PATRIOT Act* [91], lei antiterrorismo aprovada em 2001, autoriza o Procurador Geral dos Estados Unidos a solicitar judicialmente o acesso às informações individualmente identificáveis coletadas pelo Centro caso sejam relevantes a investigações relacionadas ao terrorismo nacional ou internacional. Como consequência, tornou-se necessário informar aos indivíduos entrevistados sobre a possibilidade de que as informações fornecidas possam ser utilizadas em formato identificável caso exigido por lei.

Ainda assim, práticas para a limitação da divulgação de dados estatísticos constam nos *Padrões Estatísticos do Centro Nacional de Estatísticas da Educação*, revisados em 2012 [94].

Dados publicados Conforme o Padrão 7-1 do Padrões Estatísticos do Centro Nacional de Estatísticas da Educação [94], citado acima, o NCES disponibiliza microdados de acesso restrito para agentes autorizados e publica microdados de uso público produzidos a partir dos primeiros. Além disso, o Escritório disponibiliza ferramentas de análise *on-line*, nas quais os microdados são agregados para a produção de estimativas tabulares ou, em alguns casos, para análises de regressão.

De acordo com o Padrão 4-2-7, que trata especificamente das ferramentas de análise *on-line*, caso os microdados acessados por essas ferramentas sejam de uso público, os

[2] https://nces.ed.gov/
[3] https://www.ed.gov/

mesmos devem ser submetidos a técnicas de perturbação para limitação de divulgação. Além disso, caso os microdados acessados por essas ferramentas sejam de acesso restrito, os mesmos devem cumprir as seguintes condições:

- o tamanho exato da amostra não pode ser divulgado;
- as edições de confidencialidade devem ter sido aprovadas pelo DRB responsável; e
- a ferramenta *on-line* deve divulgar apenas contagens ponderadas.

Finalmente, o Padrão 4-2-8 define que todos os microdados de uso público devem ser tratados pelo NCES de modo a limitar a possibilidade de que indivíduos venham a ser reidentificados (i.e., tornar-se *outliers* ou únicos) com o auxílio de informações externas à publicação. Dessa forma, os microdados de uso público devem passar por uma análise de risco de divulgação em preparação para a liberação, o que pode incluir técnicas de perturbação ou a redução da precisão dos dados divulgados.

2.2.4 Estudo de caso internacional 3: Divulgação de dados estatísticos pelo Escritório do Censo

Na Seção 2.2.2, cobrimos o panorama geral das agências federais dos EUA; nesta seção, dedicamos especial atenção a outra destas agências de particular interesse no escopo deste projeto: o Escritório do Censo (*United States Census Bureau* - USCB). [4]. Como uma agência estatística do Departamento de Comércio dos Estados Unidos (*United States Department of Commerce*), [5] o USCB é a maior agência estatística do governo federal. Dentre suas responsabilidades, está a condução dos Censos Decenais, requisitados pela Constituição dos EUA, e cujos resultados servem de base para a repartição dos assentos na Câmara dos Deputados entre os estados, além da distribuição de mais de US$ 675 bilhões em fundos federais para estados e organizações locais.

Além disso, o USCB é autorizado a conduzir pesquisas para outras agências federais, como regulamentado pelos Títulos 13 e 15 do Código dos Estados Unidos. Entretanto, as informações coletadas sob o mandato do Título 13, como aquelas referentes aos Censos Decenais, são protegidas por garantias de confidencialidade mais restritivas.

Regulamentação específica O Título 13 do Código dos Estados Unidos [88], em sua seção 9, determina que toda informação obtida com base nesse Título deve ser utilizada exclusivamente para os propósitos estatísticos para os quais foi coletada, sendo que nenhum indivíduo ou estabelecimento em particular pode ter sua identidade revelada a partir das informações prestadas.

[4] https://www.census.gov
[5] https://www.commerce.gov/

Para tanto, o Escritório determina práticas para a limitação da divulgação de dados estatísticos em seus *Padrões de Qualidade Estatística*, revisados em 2013 [22]. De acordo com o Requerimento S1-2, técnicas de prevenção devem ser utilizadas para impedir a liberação de informações protegidas, particularmente aquelas referentes à identificação de pessoas ou estabelecimentos, sendo que as técnicas citadas são semelhantes àquelas apresentadas pelo *Documento de Trabalho sobre Política Estatística 22* [93], detalhadas anteriormente.

Dados publicados Conforme o Requerimento A1-2 dos *Padrões de Qualidade Estatística* do USCB [22] e o Documento 22 [93], o USCB publica tanto dados agregados na forma de tabelas de frequência ou de magnitude, quanto microdados. Especificamente para os microdados, a divulgação pode ocorrer na forma de [25]:

- *Arquivos de Uso Público* (*Public Use Files - PUFs*), os quais contêm todos os registros de todos os entrevistados. Este tipo de arquivo não é publicado para os dois principais e mais abrangentes estudos realizados pelo USCB: os Censos Decenais e as Pesquisas de Comunidades Americanas (*American Community Surveys - ACS*).
- *Amostras de Microdados de Uso Público* (*Public Use Microdata Samples - PUMS*), os quais contêm os registros para apenas uma amostra da população. Este tipo de arquivo é usado, em particular, para os Censos Decenais.
- *Acesso restrito via salas seguras (Research Data Centers - RDCs)*: paralelamente à aplicação de técnicas de limitação de divulgação aos dados publicados, permite-se o acesso controlado e limitado aos microdados originais por pesquisadores e funcionários previamente autorizados. Estas salas estão distribuídas pelo país, e todas as análises devem ser realizadas dentro do ambiente regulado, o que permite ao USCB garantir a segurança dos registros e a aplicação de métodos de limitação de divulgação aos resultados obtidos antes de sua liberação para o público.

No estudo de caso do USCB, focaremos nas publicações referentes aos Censos Decenais, devido a sua importância legal, à abrangência de indivíduos entrevistados, e aos desenvolvimentos recentes nas técnicas de proteção à privacidade adotadas pelo USCB com o objetivo de garantir o direito à privacidade definido no Título 13 do Código dos Estados Unidos [88].

2.2.4.1 Técnicas utilizadas nos Censos de 1960 a 2010 para a publicação de microdados na forma de arquivos PUMS

De 1960 a 2010, diferentes técnicas de limitação de divulgação foram utilizadas nos Censos Decenais, muitas vezes motivadas por alterações nos questionários utilizados pelo Escritório do Censo [24]. A Tabela 2.2.2, apresenta um resumo das práticas adotadas pelo USCB nesse período. Estas práticas são descritas em mais detalhes a seguir.

Ano do Censo	Práticas adotadas
1960	✗
1970	✗
1980	❶
1990	❶ ❷
2000	❶ ❸ ❹ ❺
2010	
... Residências	❶ ❸ ❹ ❺
... Quartos de grupos	❶ ❹ ❻

Tabela 2.2.2: Práticas de proteção à privacidade adotadas pelo USCB nos Censos de 1960 a 2010 para a publicação de Amostras de Microdados de Uso Público (PUMS), conforme [24]: ❶ truncamento e recodificação, ❷ apagamento e imputação, ❸ permutação, ❹ truncamento de variáveis categóricas, ❺ adição de ruído, ❻ dado parcialmente sintético.

Nos primeiros dois Censos Decenais relatados, em 1960 e 1970, foi realizada apenas a remoção de identificadores diretos e a limitação do detalhamento geográfico a populações de 250.000 pessoas, técnicas que continuaram em uso até o Censo de 2010. Já no Censo de 1980, o detalhamento geográfico aumentou com a alteração da limitação a populações de 100.000 pessoas, mas também foram introduzidos agrupamentos de renda e truncamento superior para as variáveis renda e idade.

Para o Censo Decenal de 1990, mais variáveis foram submetidas à recodificação e ao truncamento superior, dentre elas: locais de residência e de trabalho, tipo de dormitório coletivo (como hospitais, prisões, e dormitórios universitários e militares), renda, e idade. Além disso, com o objetivo de proteger tabelas publicadas para pequenas populações, a técnica de apagamento e imputação também foi aplicada aos PUMS.

Para o Censo do ano 2000, a técnica de apagamento e imputação deixou de ser utilizada, mas a recodificação e o truncamento superior de variáveis foram mantidos, agora acrescentado o truncamento inferior. Uma vez que truncamentos superiores deveriam ter ao menos três valores, a maioria das variáveis econômicas foram truncadas superiormente a nível nacional. Além disso, utilizou-se pela primeira vez da permutação de registros para pequenas regiões geográficas, do truncamento de variáveis categóricas, e da adição de ruído para residências grandes.

Finalmente, para o Censo de 2010, técnicas diferentes foram utilizadas para residências e quartos de grupos, uma vez que permutações ou adição de ruído não geram resultados aceitáveis para quartos de grupos devido à homogeneidade normalmente detectada nessas populações. Dessa forma, optou-se pela adição de dados sintéticos como técnica de limitação de divulgação para quartos de grupos, enquanto as técnicas aplicadas no Censo do ano 2000 foram mantidas para o caso das residências [21].

2.2.4.2 Técnicas utilizadas no Censo de 2020

O USCB está atualmente executando o Censo de 2020, no qual ocorrerá a maior alteração já realizada pela agência no que diz respeito às suas práticas de limitação de divulgação de dados. As mudanças são devidas ao reconhecimento de que os métodos utilizados até o Censo Decenal de 2010, conhecidos como métodos sintáticos, [6] são inerentemente inseguros e sujeitam os registros individuais à possibilidade de reidentificação [25]. Mais precisamente:

- Para o caso de microdados, a reidentificação pode ser realizada, e.g., determinando-se combinações únicas de variáveis (quaseidentificadores) e combinando-as com informações disponíveis externamente, como em outras publicações de dados estatísticos ou mesmo na Internet.
- Para o caso de dados tabulares, [7] a reidentificação pode ser realizada, e.g., ao vincularem-se tabelas publicadas pelo mesmo estudo estatístico em busca de células que possam ser ligadas de modo a recriar os microdados originais [25]. Outro risco é o *ataque de reconstrução de bases de dados*, descoberto por Nissim em 2003 [36]. Seu famoso *Teorema da Reconstrução de Bases de Dados* demonstrou que mesmo uma pequena quantidade de consultas a uma base de microdados (ou divulgação de dados tabulares) seria suficiente para revelar informações de registros únicos. Desde então, o USCB já realizou estudos de reidentificação em seus microdados e encontrou combinações de variáveis que poderiam ser utilizadas em ataques de reidentificação. Entretanto, esses estudos não podem ser publicados devido à regulamentação que rege a confidencialidade dos dados obtidos pelo Escritório [57].

Dessa forma, os métodos sintáticos não serão utilizados no Censo Decenal de 2020, dando lugar aos métodos semânticos, [8] mais especificamente à privacidade diferencial (*differential privacy*, no original em inglês), inicialmente proposta por Dwork et al. em 2006 [51]. A privacidade diferencial garante formalmente que apenas uma quantidade negligível de informação sobre um dado indivíduo pode ser obtida a partir dos dados publicados, independemente da participação ou não do indivíduo na pesquisa [50].

Entretanto, a implementação da privacidade diferencial pelo USCB enfrentou e enfrenta diversas dificuldades, tanto técnicas [58] quanto de comunicação com o restante da comunidade de estatística [73]. Apesar disso, o Escritório do Censo continua a preparação para o uso da privacidade diferencial, inclusive com a publicação de produtos de demonstração derivados dos microdados do Censo de 2010 [23].

[6] *Métodos sintáticos* são aqueles que formulam condições puramente sintáticas para a divulgação de informação. Exemplos incluem as técnicas de desidentificação, de *k-anonymity* e de *l-diversity*.

[7] É importante ressaltar que no presente contexto das agências federais dos EUA, o termo *dados tabulares* é utilizado como sinônimo de *dados agregados*, e não se confunde com *microdados* [93].

[8] *Métodos semânticos* são aqueles que formulam condições semânticas para a divulgação de informação, incorporando em si mesmos a formalização de privacidade almejada.

Devido às alterações nos métodos utilizados pelo USCB para o controle da divulgação de dados estatísticos, mudanças também são esperadas no que diz respeito aos produtos que serão disponibilizados pelo Escritório a partir dos dados coletados pelo Censo de 2020. [9] Até a última atualização divulgada pelo USCB, em 08 de novembro de 2019, por volta de um terço das publicações agregadas tabulares planejadas já são suportadas pelo novo sistema baseado em privacidade diferencial, ou estão em fase de implementação. Entretanto, os outros dois terços das publicações agregadas tabulares ainda estão em fase de pesquisa para adaptação, ou por serem dependentes de operações complexas, ou por serem muito específicas no que se refere à população abordada, ou ainda por dependerem de publicações ainda não suportadas. Finalmente, ainda não há, também, uma definição quanto à publicação ou não de arquivos PUMS referentes ao Censo de 2020 [26].

2.3 Contexto da União Europeia

Nesta seção, após um sumário dos aspectos mais relevantes da regulamentação geral da União Europeia sobre privacidade, apresentamos um estudo de caso relativo à divulgação de dados educacionais por escolas e pelo Ministério da Educação da Holanda.

2.3.1 Regulamentação geral

Em 1995, o Parlamento e o Conselho Europeu acataram a Diretiva 95/46/CE, relativa à proteção dos indivíduos no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados [28].

Em 25 de maio de 2018, entrou em vigor o Regulamento UE 2016/679 [29], conhecido como *Regulamento Geral de Proteção de Dados* - RGPD em Portugal, e no Brasil por seu nome e sigla em inglês, *General Data Protection Regulation* - GDPR. O regulamento trata sobre privacidade e proteção de dados pessoais de todos os indivíduos na União Europeia (UE) e no Espaço Econômico Europeu (EEE) –incluindo entidades que realizam comércio com a UE, independentemente de seu país de origem–, assim como da exportação de dados pessoais para fora dessas regiões. O regulamento tem como objetivo prover aos cidadãos e residentes formas de controlar seus dados pessoais, além de unificar o quadro regulamentar europeu, baseando-se nos seguintes princípios:

[9] De acordo com [58], os dados coletados pelo USCB para o Censo de 2020 proveem da Internet, de chamadas telefônicas e de formulários entregues por funcionários da agência na residência de cada cidadão. Todos esses dados confidenciais são coletados e processados para formar a primeira base de dados não editada, chamada de Arquivo Não Editado do Censo (*Census Unedited File* - CUF). Especialistas de diversas áreas realizam, então, correções de problemas de diferentes naturezas no CUF utilizando como auxílio outras fontes de dados, e a partir disso é gerada a primeira base editada, chamada de Arquivo Editado do Censo (*Census Edited File* - CEF). Em seguida, algumas modificações são realizadas no CEF pelo Sistema de Prevenção de Divulgação (*Disclosure Avoidance System* - DAS), e o resultado destas modificações compõe a base chamada de Arquivo Detalhado de Microdados (*Microdata Detailed File* - MDF). Todos os dados disponibilizados publicamente são obtidos utilizando a MDF.

- **transparência:** o titular dos dados processados deve estar ciente disso, ter dado sua permissão para isso, e conhecer seus direitos;
- **limitação de finalidade:** os dados pessoais são coletados para uma finalidade legítima específica e não podem ser usados para outras finalidades;
- **limitação de dados:** somente os dados necessários para a finalidade pretendida podem ser coletados;
- **precisão:** os dados pessoais devem estar e permanecer corretos;
- **limite de retenção:** os dados pessoais não podem ser mantidos por mais tempo que o necessário para a finalidade pretendida;
- **integridade e confidencialidade:** os dados pessoais devem ser protegidos contra acesso, perda ou destruição não autorizada; e
- **responsabilidade:** o responsável pelo tratamento deve demonstrar conformidade com essas regras.

Os principais direitos que a GDPR garante aos cidadãos europeus são sumarizados por Filipe Pontes como a seguir [99].

1. **Direito de informação:** direito de receber informações sobre os termos do tratamento de dados pessoais quando da sua recolha;
2. **Direito de acesso**: direito de obter confirmação de que os dados pessoais são ou não objeto de tratamento e, se for o caso, ter acesso aos seus dados pessoais;
3. **Direito de retificação**: o titular tem o direito de obter, sem demora injustificada, a retificação ou atualização dos dados pessoais inexatos;
4. **Direito de apagamento dos dados**: o titular tem o direito de obter o apagamento dos seus dados pessoais, sem demora injustificada, dentro dos limites legalmente previstos;
5. **Direito à limitação de tratamento**: o titular dos dados tem o direito de obter a limitação do tratamento;
6. **Direito de Portabilidade**: o titular dos dados tem o direito de receber os dados pessoais que lhe digam respeito e que tenha fornecido num formato de uso corrente e de leitura automática;
7. **Direito de não ficar sujeito a decisões individuais automatizadas**: o titular dos dados tem o direito de não ficar sujeito a nenhuma decisão tomada exclusivamente com base no tratamento automatizado dos seus dados pessoais, incluindo a definição de perfis, que produza efeitos na sua esfera jurídica ou que o afete significativamente de forma similar;

8. **Direito de oposição:** o titular dos dados tem o direito de se opor, a qualquer momento, por motivos relacionados com a sua situação particular, ao tratamento dos dados pessoais que lhe digam respeito, nomeadamente quando os seus dados sejam tratados para efeitos de comercialização direta.

2.3.2 Estudo de caso internacional 4: Divulgação de dados educacionais por escolas e pelo Ministério da Educação da Holanda

A variedade de países constituintes da União Europeia permite uma vasta gama de opções para a escolha do estudo de caso europeu apresentado no presente documento. Os seguintes três motivos nos levaram a optar por um caso da Holanda.

Como primeiro motivo, um dos autores deste documento é de nacionalidade holandesa e, como pesquisador experiente na área de segurança, tem conhecimento profundo sobre as questões de privacidade de seu país.

Como segundo motivo, a sociedade holandesa, por razões históricas, sempre demonstrou um nível de preocupação de vanguarda com questões relativas à privacidade de seus cidadãos. Podemos citar os seguintes exemplos que ilustram este notório histórico.

- Durante a ocupação pela Alemanha na Segunda Guerra Mundial, os nazistas usaram cadastros dos cidadãos (na Holanda o papel dos cartórios é obrigação dos municípios) para perseguir os judeus, e a Resistência Holandesa destruiu o cadastro da cidade Amsterdam para sabotar essa política [1].
- A Holanda realiza um censo (*volkstelling*) estatístico populacional desde 1795. No entanto, o censo de 1981 foi cancelado quando ficou evidente que 26% da população se recusaria a colaborar por motivos de privacidade. Em 1991, o censo foi oficialmente abolido e substituído por técnicas por amostragem [5].
- Durante a atual pandemia de COVID-19, o governo holandês tem tentado empregar um aplicativo rastreador (*tracing app*, em inglês) para identificar indivíduos com quem uma pessoa possivelmente infectada teve contato, facilitando assim o rastreamento do espalhamento da doença. No entanto, há em curso um grande debate na Holanda sobre como essa medida poderia ser implementada, com o objetivo de mitigar a séria erosão de privacidade que ela poderia causar. De fato, a equivalente holandesa à nossa Autoridade Nacional de Proteção de Dados já demonstrou independência ao contrariar várias vezes o governo [108].

Como terceiro motivo para a opção por um caso holandês, notamos que a tecnologia se integrou ao sistema educacional do país precocemente, e sistemas computacionais passaram a fazer parte da administração das escolas há pelo menos três décadas. Cada escola na Holanda tem um sistema de administração que armazena dados pessoais de alunos

e professores, e existe também um sistema de acompanhamento escolar (*leerlingvolgsysteem*), para acompanhar o progresso acadêmico do aluno. Além disso, com o advento da Internet, o computador está cada vez mais entremeado nas salas de aula: por um lado as editoras de livros didáticos oferecem cada vez mais material *on-line*, enquanto várias empresas oferecem ferramentas de ensino *on-line*.

Desta forma, formou-se na Holanda um ambiente bastante complexo com diversas partes interessadas em dados educacionais: as escolas, as empresas que desenvolvem sistemas computacionais para as escolas, as empresas que oferecem os serviços de TI, as editoras de material didático, o Ministério da Educação, entre outros. Além disso, há vários tipos de usuários de informação, incluindo professores, alunos, diretores, coordenadores, e administradores. Por causa deste contexto, a Holanda teve que desenvolver de forma pioneira suas próprias políticas sobre quem pode ter acesso a que tipo de informação, e sobre as garantias de proteção da privacidade de dados pessoais das partes envolvidas.

Como um exemplo desta forte atuação holandesa nessa questão, podemos citar um relevante caso ocorrido em 2018. Naquele ano, a Autoridade de Privacidade - AP (*Autoriteit Persoonsgegevens*), equivalente à Autoridade Nacional de Proteção de Dados (ANPD) brasileira, conduziu uma investigação sobre privacidade educacional envolvendo a ASKO (*Amsterdamse Stichting voor Katholiek, Protestants-Christelijk en Interconfessioneel Onderwijs*), um conselho que administra 32 escolas na Holanda, e que empregava um sistema computacional cujo nome foi omitido para proteger o fornecedor, sendo chamado apenas de "Sistema de Rastreamento de Estudantes X". Os resultados desta investigação foram publicados em um relatório intitulado *Pesquisa sobre os direitos de acesso dos funcionários das instituições de ensino da ASKO a dados pessoais de estudantes no Sistema de Rastreamento de Estudantes X* [15]. A motivação dada a esse esforço foi a seguinte:

> As instituições de ensino processam muitos dados pessoais dos alunos, como detalhes de contato, número de previdência social (BSN), dados de ausência às aulas, resultados de estudos e dados sobre saúde e bem-estar. Elas precisam dessas informações para fornecer educação aos alunos. Tendo em vista a natureza e o escopo desses dados pessoais, incluindo dados pessoais especiais, e o fato de as crianças poderem ser classificadas como pessoas vulneráveis, é de grande importância que as instituições educacionais manejem com cuidado os dados pessoais e os princípios da Lei de Proteção de Dados Pessoais e observem os princípios do Regulamento Geral de Proteção de Dados (GDPR) a partir de 25 de maio de 2018. A Autoridade Holandesa de Proteção de Dados (AP), portanto, anunciou em sua agenda de 2016 que será dada especial atenção à privacidade das crianças na educação.

O relatório final da investigação chegou a definir recomendações, diretrizes e boas práticas a serem adotadas no tratamento de dados educacionais. Entretanto, nenhuma multa foi aplicada à época uma vez que o relatório foi publicado em janeiro de 2018, antes do GDPR entrar em vigor. Como reação a esse relatório, e para agir de acordo com

a GDPR, foi criado um consórcio que combinou com todas as partes que atuam no ambiente educacional um acordo de privacidade educacional [101]. Discutimos abaixo as ações deste consórcio no contexto da regulamentação específica sobre privacidade nos dados educacionais da Holanda.

Regulamentação específica Já em 1989 foi aprovada na Holanda a *Lei Registro de Pessoas* (*Wet persoonsregistraties*) [7], determinando que um indivíduo precisaria conceder permissão a uma empresa ou instância no compartilhamento de seus dados para uso diferente do objetivo original. A lei também criou uma Câmara de Privacidade para cadastrar esses registros e para se certificar do cumprimento dos regulamentos.

Em setembro de 2001, entrou em vigor a *Lei de Proteção a Dados Pessoais (Wet bescherming persoonsgegevens)* [6] para proteger a privacidade dos cidadãos, substituindo a lei anterior. Essa lei foi em grande parte baseada na Diretiva 95/46 da Comunidade Europeia relativa à proteção de dados pessoais [28]. A lei concedia aos cidadãos certos direitos, como o direito de saber o que está acontecendo com seus dados pessoais.

Em maio de 2018, essa lei foi substituída pela GDPR europeia e, portanto, a partir de então a Holanda passou a seguir o regulamento europeu discutido na seção anterior. A lei holandesa que implementa a GDPR é conhecida como *Algemene Verordening Gegevensbescherming* (AVG) [4]. A Câmara de Privacidade passou a chamar-se *Autoridade de Privacidade* - AP (*Autoriteit Persoonsgegevens*), [10] o equivalente holandês da Autoridade Nacional de Proteção de Dados prevista pela LGPD brasileira.

No contexto de tratamento de dados educacionais, a AP holandesa provê instruções especialmente cuidadosas, mesmo por que os indivíduos a quem os dados se referem são muitas vezes crianças, que têm direito a proteção e tratamento especiais pela GDPR. A AP esclarece precisamente várias questões relativas à privacidade na educação [14], incluindo pontos como os seguintes:

- **Responsabilização**: as escolas devem ser capazes de demonstrar quais as medidas técnicas e organizacionais tomadas para proteger os dados pessoais dos alunos.
- **Necessidade de aprovação**: caso uma escola deseje estabelecer ou alterar um regulamento para o uso de dados pessoais, ela deve primeiro pedir aprovação ao conselho de participação.
- **Câmeras nas escolas**: entre outros, oferece "Vigilância por câmera na escola", um pacote gratuito de lições sobre privacidade adaptado para alunos de 11-12 anos.

Além disso, a AP também estabelece claramente obrigações das partes envolvidas no tratamento de dados escolares. Alguns exemplos de questões detalhadas pela AP incluem:

- **Obrigações das escolas em relação ao GDPR**:

[10] https://autoriteitpersoonsgegevens.nl

  - Quando cada escola deve realizar uma avaliação de impacto de proteção de dados (*Data Protection Impact Assessment* - DPIA).
  - Quais regras se aplicam ao usar um sistema de rastreamento de alunos.
  - Quais regras se aplicam em relação ao programa de saúde do município para os alunos.

- **Obrigações das escolas sobre informações e permissões**:
  - Quando cada escola deve informar alunos e pais sobre os dados pessoais por ela processados.
  - Quais informações cada escola deve fornecer aos alunos e/ou pais sobre os dados pessoais que processa.
  - Quais cuidados tomar ao informar pais e alunos sobre os dados processados.
  - Como e se cada escola pode publicar imagens de alunos.

- **Obrigações das escolas sobre programas de ação em que dados pessoais são utilizados**:
  - Quem é responsável pelo processamento de dados ao usar um programa anti-*bullying*.
  - Como e se cada escola pode incluir informações sobre a raça, religião ou orientação sexual de uma pessoa em um programa anti-*bullying*.
  - Se e quando cada escola precisa de uma base legal para processar os dados pessoais de alunos.
  - Por quanto tempo cada escola pode manter os dados dos alunos.
  - Como cada escola deve gerenciar a segurança ao usar um programa anti-*bullying*.
  - Quem pode acessar os dados dos alunos.

Como mencionado anteriormente, há em vigência na Holanda um acordo de privacidade educacional criado por um consórcio formado por todas as partes que atuam no ambiente das escolas [101]. Um dos pontos principais desde acordo é a definição de um código único para cada aluno de ensino básico e secundário, que serve como pseudônimo (o que consiste em uma técnica de *pseudonimização*). O consórcio também é responsável pela definição técnica dessa proposta. Foi estabelecido um padrão que define protocolos de Internet [52], de como as partes devem trocar as informações entre si [69], e inclui uma implementação de referência [68].

Dados divulgados por escolas A partir da inspeção da vasta documentação disponível, concluímos que as escolas holandesas nunca publicam microdados dos alunos. Além disso, uma pessoa só tem acesso aos dados realmente necessários para uma dada

tarefa que esteja executando (*need-to-know principle*, em inglês). Como um exemplo, normalmente até funcionários das escolas não têm acesso aos dados pessoais do aluno, e quando um aluno precisa de assistência imediata emergencial, um funcionário entra num procedimento específico que lhe concede acesso aos dados do aluno durante 24 horas. Esse acesso fica registrado, e o funcionário precisa justificá-lo.

Dados divulgados pelo Ministério da Educação Na Holanda, a Divisão de Implantação Ensino (*Dienst Uitvoering Onderwijs*) do Ministério da Educação é responsável por gerir as escolas e universidades do país. A informação divulgada por esta divisão está disponível para visualização *on-line* [35]. Nunca são divulgados microdados de alunos, sob nenhuma forma. São divulgados apenas dados de escolas ou regiões, e dados de alunos apenas em forma agregada. Exemplos de informações disponibilizadas incluem:

- endereços das instituições de ensino;
- dados financeiros das instituições de ensino;
- quantidade de alunos e funcionários por instituição de ensino;
- indicadores de qualidade por instituição de ensino;
- distância média entre a casa do aluno e sua escola;
- dados sobre a porcentagem de alunos que abandonaram um curso de uma escola antecipadamente, sem se formar;
- resultados médios das escolas para exames nacionais; e
- previsão de quantidade de alunos por tipo de ensino e por região.

2.4 Contexto da Austrália

Nesta seção, após um sumário dos aspectos mais relevantes da regulamentação geral australiana sobre privacidade, apresentamos um estudo de caso no contexto da Austrália.

2.4.1 Regulamentação geral

A *Lei de Privacidade* de 1988 [81], revisada em novembro de 2015, estabelece a regulação nacional para privacidade e tratamento de informações pessoais, particularmente com a definição dos *Princípios Australianos de Privacidade*.

Dentre os princípios estabelecidos, destacam-se a garantia do uso de *anonimização*, i.e., a desassociação dos indivíduos de seus respectivos registros, ou de *pseudonimização*, i.e., a atribuição de um código individual a cada registro de uma base de dados, além da transparência por parte das entidades governamentais na coleta e no tratamento de informações pessoais, assim como de regras para o uso e a divulgação de dados pessoais.

2.4.2 Estudo de caso internacional 5: Divulgação de dados estatísticos pelo Escritório Australiano de Estatística

O Escritório Australiano de Estatística (*Australian Bureau of Statistics* - ABS) foi criado por Lei em 1975 [79] e tem seu mandato de coleta de dados estatísticos estabelecido pela Lei do Censo e Estatísticas de 1905 [80], o que inclui a realização do Censo Nacional. A Lei de 1905 requer que o ABS publique as informações estatísticas ao mesmo tempo em que garanta a confidencialidade das informações coletadas.

Regulamentação específica As informações pessoais coletadas para o Censo são regidas pela *Política de Privacidade* do ABS [82, 86]. De acordo com a Legislação vigente e a *Política de Privacidade*, o Escritório nunca publicou, nem publicará, informações identificáveis sobre indivíduos, residências, ou estabelecimentos comerciais.

Assim como a Nova Zelândia e o Canadá [86], o Escritório armazena nomes e endereços dos entrevistados, mas de forma desassociada entre si e das demais informações coletadas, sendo utilizadas apenas para a reconstrução da base de dados quando necessário.

Dados divulgados Até o Censo de 2016, o ABS disponibilizou microdados apenas por amostras da população e sob restrições de acesso [84]. Além do controle de acesso aos arquivos de amostras, o Escritório também cobra pelo serviço de acordo com o acesso requisitado [85].

Paralelamente, é mantido o *Australian Census Longitudinal Dataset* (ACLD) [83], um Censo Longitudinal criado a partir de amostras aleatórias com correspondência, a saber:

- 2006-2011 ACLD: amostra de 5% de registros aleatórios, o equivalente a aproximadamente um milhão de registros, do Censo de 2006 com os registros correspondentes do Censo de 2011.
- 2011-2016 ACLD: aproximadamente 1,2 milhão de registros do Censo de 2011 com os registros correspondentes do Censo de 2016.

O Censo Longitudinal é acessível para usuários registrados no serviço de geração de tabelas e na forma de microdados para usuários aprovados que estejam conduzindo projetos consentidos. Assim como ocorre para os microdados em Censos individuais, o Escritório também cobra pelos serviços do Censo Longitudinal.

Finalmente, vale destacar que nenhuma informação sobre a aplicação de métodos de anonimização foi encontrada para as divulgações estatísticas do Escritório Australiano de Estatística. As únicas medidas de controle de divulgação relatadas foram a remoção de identificadores óbvios e a publicação de amostras.

Controvérsia sobre armazenamento de dados pessoais do Censo de 2016 Alterando as práticas até então empregadas, o Escritório Australiano de Estatística decidiu manter nomes e endereços coletados para o Censo de População e Habitação de 2016. Em comunicado oficial [11], o ABS justificou que a medida permitiria uma imagem estatística mais rica e dinâmica da Austrália por meio da combinação de dados do Censo com outros dados administrativos e de pesquisa. Mais especificamente, o Escritório declarou que:

> Embora o Censo sempre tenha sido valioso por si só, quando usado em combinação com outros dados, ele pode fornecer uma visão ainda maior. Alguns exemplos incluem:
>
> - A combinação de dados do Censo com dados educacionais pode fornecer informações sobre os resultados de emprego a partir dos vários caminhos educacionais disponíveis para os australianos; e
> - A combinação de dados do Censo com dados de saúde pode ajudar a melhorar a compreensão e o apoio da Austrália de pessoas que necessitem de serviços de saúde mental e ajudar no projeto de melhores programas de apoio e prevenção.

O Escritório alegou ainda que a retenção de endereços permitiria operações de pesquisa mais eficientes, reduzindo o custo para os contribuintes e o ônus para as famílias australianas.

No mesmo comunicado oficial, o ABS afirmou comprometimento com a proteção da privacidade e confidencialidade dos respondentes ao Censo. Declarou ainda que a medida se baseou em uma cuidadosa Avaliação de Impacto de Privacidade realizada em 2015 [10], e que os riscos à privacidade identificados seriam considerados pequenos dadas as medidas de proteção que o Escritório já adotava. De acordo com o Escritório, tais riscos seriam ainda atenuados pelo armazenamento de nomes e endereços separadamente de outros dados do Censo, bem como separadamente um do outro. Entretanto, houve reação contundente por vários setores da sociedade australiana contra medida.

Em particular, houve forte resistência por parte da Fundação Australiana de Privacidade (*Australian Privacy Foundation*), [11] uma organização não governamental formada com o objetivo de proteger os direitos de privacidade dos australianos. O vice-diretor da entidade, que tem como uma de suas missões defender o direito dos indivíduos de controlar o acesso às informações pessoais e de evitar invasões excessiva, declarou à época que [12]:

> As mudanças na tecnologia aumentaram o risco [à privacidade], de modo que a decisão do ABS de armazenar nossos nomes e endereços por anos é ainda mais preocupante. [Esta decisão] cria um atrativo irresistível para hackers e

[11] https://privacy.org.au/

> cibercriminosos em uma época em que nenhuma segurança de tecnologia da informação pode impedir "intrusos motivados". As violações graves de dados são agora um perigo real e crescente. O armazenamento de nomes e endereços também atrairá dezenas de agências [governamentais] reclamando direito de acesso aos dados. O cruzamento de dados está mais fácil, mais barato e mais intrusivo do que nunca. Pode ser necessária apenas "uma canetada" ou uma emenda a um regulamento para que o Governo use os dados para outros fins, caso decida fazê-lo.

À época a imprensa australiana engrossou o coro às críticas à posição do ABS, notando que o movimento do Escritório contrariava as práticas adotadas até então e abalava a credibilidade do Escritório frente à população [98].

A sensação de desconforto de parte da sociedade australiana foi reforçada pela ocorrência de ataques aos servidores do ABS em 2016. Apesar de os ataques não terem causado vazamento ou comprometimento de dados dos Censos, eles reforçaram as dúvidas de parte da imprensa e da sociedade australiana sobre os riscos de vazamento dos dados pessoais armazenados pelo Escritório [17, 103]. De fato, todo o ocorrido tem exigido um grande esforço por parte do ABS para a recuperação de sua credibilidade junto à sociedade australiana.

3 Estudos de caso do panorama nacional

Nesta seção provemos um panorama nacional dos diferentes métodos pelos quais institutos produtores de estatísticas oficiais divulgam informação de forma a garantir transparência e, ao mesmo tempo, privacidade de dados pessoais. Iniciamos a seção com um sumário das principais diretivas e regulamentações gerais brasileiras sobre transparência e privacidade, com especial foco na Lei de Acesso à Informação (LAI) e na Lei Geral de Proteção de Dados Pessoais Pessoais (LGPD), além do Marco Civil da Internet (MCI). Em seguida, apresentamos estudos de caso referentes à divulgação de informação pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e pelo Portal da Transparência do Governo Federal. [1]

3.1 Regulamentação geral

Nesta seção apresentamos um sumário da regulamentação brasileira sobre privacidade e transparência, com foco nas previsões da Constituição da República, na Lei de Acesso à Informação e na Lei de Proteção de Dados Pessoais. [2]

A Constituição da República Federativa do Brasil A Constituição Brasileira de 1988 [37], em seu Artigo 5º, garante a todos os brasileiros e estrangeiros residentes no país que:

> X são invioláveis a intimidade, a vida privada, a honra e a imagem das pessoas, assegurado o direito a indenização pelo dano material ou moral decorrente de sua violação;

[1] Um panorama da percepção da questão de privacidade pela sociedade brasileira, formado por uma seleção de reportagens e artigos relevantes recentes produzidos sobre o assunto, pode ser encontrado no Anexo E.

[2] Aqui nos focamos em uma análise dos pontos mais relevantes dessas diretrizes; uma seleção textual dos pontos da Constituição Federal, da LAI e da LGPD mais relevantes ao Projeto PRICE encontram-se nos Anexos A, B e C, respectivamente.

XXXIII todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado;

Dessa forma, tem-se que o inciso X do Artigo 5º fundamenta o direito constitucional à privacidade do indivíduo, enquanto que o inciso XXXIII do mesmo Artigo fundamenta o direito constitucional, individual e coletivo, à transparência por parte do Estado. Entretanto, não há qualquer definição referente a como equilibrar esses dois princípios ou sobre quais seriam os limites legais de cada um deles, particularmente frente aos avanços tecnológicos das últimas décadas, os quais propiciaram a coleta e o tratamento em larga escala de dados pessoais.

Portanto, devido à necessidade de regulamentação dos direitos à privacidade e à transparência determinados nas Cláusulas Pétreas da Constituição de 1988, a Lei 12.527 de 2011 foi sancionada para regulamentar o acesso à informação e a Lei 13.709 de 2018 foi sancionada para regulamentar a proteção de dados pessoais.

Lei 12.527/2011 - Lei de Acesso à Informação (LAI) A Lei 12.527 de 2011 [47], conhecida como Lei de Acesso à Informação (LAI), regulamenta o inciso XXXIII do Artigo 5º da Constituição Federal. Em seu Artigo 4º, essa Lei define o significado de alguns termos técnicos, dentre eles:

I **informação**: dados, processados ou não, que podem ser utilizados para produção e transmissão de conhecimento, contidos em qualquer meio, suporte ou formato;

II **documento**: unidade de registro de informações, qualquer que seja o suporte ou formato;

III **informação sigilosa**: aquela submetida temporariamente à restrição de acesso público em razão de sua imprescindibilidade para a segurança da sociedade e do Estado;

IV **informação pessoal**: aquela relacionada à pessoa natural identificada ou identificável;

V **tratamento da informação**: conjunto de ações referentes à produção, recepção, classificação, utilização, acesso, reprodução, transporte, transmissão, distribuição, arquivamento, armazenamento, eliminação, avaliação, destinação ou controle da informação;

VI **autenticidade**: qualidade da informação que tenha sido produzida, expedida, recebida ou modificada por determinado indivíduo, equipamento ou sistema;

VII **integridade**: qualidade da informação não modificada, inclusive quanto à origem, trânsito e destino;

VIII **primariedade**: qualidade da informação coletada na fonte, com o máximo de detalhamento possível, sem modificações.

Os Artigos 6º e 8º da LAI determinam ao poder público o dever de garantir o amplo acesso à informação, particularmente àquela considerada de interesse coletivo ou geral, a qual deve ser disponibilizada via Internet independentemente de requerimento. Além disso, o Artigo 7º garante o direito de acesso às demais informações não imediatamente disponibilizadas via Internet. Ademais, os Artigos 7º e 8º garantem o acesso à informação primária, íntegra, autêntica, e atualizada em ambos os casos, conforme as definições dadas no Artigo 4º.

Entretanto, ainda de acordo com o Artigo 7º e de acordo com o Artigo 22º, o acesso à informação pode ser negado total ou parcialmente no caso de a informação ser considerada sigilosa, incluindo os casos de segredo de justiça ou industrial.

Particularmente para o caso do tratamento de informações pessoais, o Artigo 31º estabelece que as mesmas devem ser tratadas de forma transparente e com respeito às liberdades e garantias individuais, conforme o inciso X do Artigo 5º da Constituição Federal, determinando restrições ao seu acesso. Mas o parágrafo 3º do mesmo Artigo determina casos específicos nos quais não é exigido o consentimento individual para o uso de informações pessoais, incluindo quando as informações forem necessárias:

II à realização de estatísticas e pesquisas científicas de evidente interesse público ou geral, previstos em lei, sendo vedada a identificação da pessoa a que as informações se referirem;

V à proteção do interesse público e geral preponderante.

Entretanto, ainda de acordo com o Artigo 31º da LAI, em seu parágrafo 5º, fica em aberto à regulamentação posterior disposições sobre o tratamento de informações pessoais.

Lei 12.965/2014 - Marco Civil da Internet (MCI) Como já mencionado antes, o Marco Civil da Internet [59] surgiu no contexto da reação brasileira à invasão aos sistemas de comunicação da Presidência pelo governo dos EUA. Apesar de ter como foco primordial os direitos dos usuários de Internet, a Lei também trata de privacidade. Em particular, o Artigo 7º estabelece que o acesso à Internet é essencial ao exercício da cidadania, e assegura aos cidadãos os seguintes direitos:

VII não fornecimento a terceiros de seus dados pessoais, inclusive registros de conexão, e de acesso a aplicações de internet, salvo mediante consentimento livre, expresso e informado ou nas hipóteses previstas em lei;

VIII informações claras e completas sobre coleta, uso, armazenamento, tratamento e proteção de seus dados pessoais, que somente poderão ser utilizados para finalidades que:

a) justifiquem sua coleta;

b) não sejam vedadas pela legislação; e

c) estejam especificadas nos contratos de prestação de serviços ou em termos de uso de aplicações de internet;

IX consentimento expresso sobre coleta, uso, armazenamento e tratamento de dados pessoais, que deverá ocorrer de forma destacada das demais cláusulas contratuais;

X exclusão definitiva dos dados pessoais que tiver fornecido a determinada aplicação de internet, a seu requerimento, ao término da relação entre as partes, ressalvadas as hipóteses de guarda obrigatória de registros previstas nesta Lei;

No contexto do presente projeto, entretanto, fica evidente que o MCI é menos relevante que outras regulamentações nacionais.

Lei 13.709/2018 - Lei de Proteção de Dados Pessoais (LGPDP ou LGPD) A regulamentação sobre o tratamento de informações pessoais ausente na LAI veio com a Lei 13.709/2018 [41], conhecida como Lei Geral de Proteção de Dados Pessoais (LGPDP ou LGPD), a qual tem por objetivo a proteção dos direitos fundamentais de liberdade e de privacidade, conforme o seu Artigo 1º. Em seu Artigo 5º, essa Lei define o significado de alguns termos técnicos, dentre eles:

I **dado pessoal**: informação relacionada a pessoa natural identificada ou identificável;

II **dado pessoal sensível**: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural;

III **dado anonimizado**: dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento;

V **titular**: pessoa natural a quem se referem os dados pessoais que são objeto de tratamento;

VI **controlador**: pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados pessoais;

VII **operador**: pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador;

IX **agentes de tratamento**: o controlador e o operador;

X **tratamento**: toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração;

XI **anonimização**: utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo;

XVIII **órgão de pesquisa**: órgão ou entidade da administração pública direta ou indireta ou pessoa jurídica de direito privado sem fins lucrativos legalmente constituída sob as leis brasileiras, com sede e foro no País, que inclua em sua missão institucional ou em seu objetivo social ou estatutário a pesquisa básica ou aplicada de caráter histórico, científico, tecnológico ou estatístico;

O Artigo 7º dispõe sobre as hipóteses nas quais o tratamento de dados pessoais é permitido, sendo que o Artigo 11º aplica-se especificamente aos dados pessoais sensíveis, conforme definido no Artigo 5º, inciso II. Em ambos os casos, o tratamento é permitido mediante o consentimento do titular ou do responsável legal no caso de dados sensíveis. Entretanto, o consentimento não é exigido nas hipóteses em que o tratamento de dados seja indispensável para, conforme o inciso II do Artigo 11º:

a) cumprimento de obrigação legal ou regulatória pelo controlador;

b) tratamento compartilhado de dados necessários à execução, pela administração pública, de políticas públicas previstas em leis ou regulamentos;

c) realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais sensíveis;

Além disso, de acordo com o Artigo 12º da LGPD, dados anonimizados não são considerados dados pessoais, exceto quando a anonimização puder ser revertida com esforços razoáveis. Para tanto, deve-se considerar fatores objetivos como o custo e o tempo necessários para reverter o processo de anonimização dadas as tecnologias disponíveis e desconsiderando-se o uso de meios de terceiros.

Particularmente no contexto de estudos em saúde pública, faz-se o uso do termo pseudonimização, não definido no Artigo 5º, mas apenas no Artigo 13º, parágrafo 4º:

§ 4º Para os efeitos deste artigo, a **pseudonimização** é o tratamento por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, senão pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e seguro. [Ênfase nossa.]

Tecnicamente, *pseudonimização* é a atribuição de um código individual a cada registro de uma base de dados, acompanhada da transferência de identificadores diretos, como nome e números de identificação, para outra base de dados. Dessa forma, o controlador é capaz de mapear identificadores diretos em uma base de dados a registros específicos na outra base através do uso dos códigos individuais. Entretanto, diferentemente do que é afirmado pelo parágrafo 4º do Artigo 13º, pseudonimização, do ponto de vista técnico, não garante proteção contra a reidentificação de indivíduos a partir da base de dados sem identificadores diretos.

Finalmente, o Artigo 16º da LGPD determina a eliminação dos dados pessoais uma vez finalizado o tratamento para o qual seu uso foi autorizado, salvo as hipóteses específicas nas quais é permitida a sua conservação.

Nota sobre o princípio de autodeterminação informativa Um fundamento em comum na regulamentação apresentada acima é o *princípio de autodeterminação informativa*, descrito da seguinte forma por Luíza Brandão, fundadora e diretora do Instituto de Referência em Internet e Sociedade (IRIS) [3] [20]:

> Esse conceito [o princípio de autodeterminação informativa], que ganha notoriedade nas discussões europeias sobre proteção de dados, leva em conta que a lógica da economia baseada em dados gira em torno de informações construídas sobre e a partir de pessoas. Em tal cenário, a autodeterminação seria uma forma de proteção frente aos interesses de outros atores da "economia da vigilância" pela qual as pessoas podem determinar, escolher e controlar as informações (extraídas a partir de seus dados) sobre elas, seus modos de utilização e, inclusive, seu apagamento ou correção. A escolha legislativa do Marco Civil que reflete esse conceito, como em outras regulações pelo mundo, se expressa na ideia de consentimento. Conforme descreve Bruno Bioni [18], o consentimento *[n]ada mais é do que a liberdade que todo cidadão tem de reger a sua vida, criando, modificando e extinguindo as suas relações em meio à sociedade. Essa mesma autonomia é captada e transportada para a proteção dos dados pessoais, na medida em que é o próprio cidadão quem deve governar seus dados pessoais.*

Nota sobre regulamentação específica referente a crianças No que se refere ao tratamento de dados de crianças e adolescentes, o Artigo 14º da LGPD estabelece que o mesmo deverá ser realizado com o consentimento específico e em destaque dado por pelo menos um dos pais ou pelo responsável legal. Além disso, os controladores responsáveis pelos dados devem publicar os tipos de dados coletados, como esses dados são utilizados, e quais são os meios para que os direitos garantidos ao titular dos dados, de acordo com a LGPD, sejam exercidos, incluindo o acesso aos dados, a correção dos mesmos, ou

[3] http://irisbh.com.br

a revogação do consentimento e eliminação dos dados. Ademais, todas as informações a respeito do tratamento de dados de crianças e adolescentes devem ser fornecidas de maneira simples, clara e acessível, de modo que seja adequada ao entendimento não apenas dos pais ou do responsável legal, mas também da criança ou do adolescente.

3.2 Estudo de caso nacional 1: Divulgação de dados estatísticos pelo IBGE

A Fundação Instituto Brasileiro de Geografia e Estatística (IBGE), [4] vinculada ao Ministério da Economia, [5] é o principal órgão estatístico do país e responsável pela produção e análise dos Censos Demográficos Decenais e de diversos outros produtos estatísticos nos âmbitos social, demográfico, agropecuário, e econômico.

3.2.1 Regulamentação específica

A Lei 5.534, de novembro de 1968 [43], junto à sua regulamentação pelo Decreto 73.177, de novembro de 1973 [40], define o mandato do IBGE para coleta de dados sob a jurisdição brasileira. De acordo com a Lei, em seu Artigo 1º, toda pessoa natural ou jurídica, de direito público ou privado, é obrigada a prestar as informações solicitadas pelo IBGE. Entretanto, essas informações são protegidas por sigilo e podem ser usadas, exclusivamente, para fins estatísticos [34].

Tendo em vista o mandato legal do IBGE e o *Código Regional de Boas Práticas das Estatísticas para a América Latina e o Caribe* [27], baseado nos *Princípios Fundamentais das Estatísticas Oficiais da ONU* [30], o IBGE publicou em 2013 o *Código de Boas Práticas das Estatísticas do IBGE* [33]. De acordo com a Seção 1 do Código, Princípio 4, o IBGE considera como indicadores de boas práticas relativas à confidencialidade estatística, que:

4.6 O acesso aos microdados não desidentificados deve estar sujeito a protocolos de confidencialidade, estabelecidos para usuários externos que têm acesso com a finalidade de análise e pesquisa estatística.

4.7 O arquivamento das informações pelo Instituto deve ser feito de acordo com os protocolos de segurança e confidencialidade estabelecidos e com as normas vigentes.

3.2.2 Dados divulgados

O IBGE divulgou em 2018 o documento *Confidencialidade no IBGE - procedimentos adotados na preservação do sigilo das informações individuais nas divulgações de resultados das operações estatísticas* [34]. Dentre os procedimentos abordados, destacam-se

[4] https://www.ibge.gov.br/
[5] https://www.gov.br/economia/pt-br

aqueles referentes às Pesquisas Domiciliares por Amostragem Probabilística e aos Censos Demográficos Decenais.

Pesquisas Domiciliares por Amostragem Probabilística No que se refere às Pesquisas Domiciliares por Amostragem Probabilística, o IBGE divulga tanto informações agregadas em forma de tabelas quanto microdados. Para os dados agregados, o Instituto afirma não aplicar procedimentos de desidentificação nas células das tabelas, sob o argumento de que a reidentificação de informantes não é possível dada a natureza da investigação e da divulgação [34]: [6]

> Isso porque a natureza da investigação e da divulgação não permite a identificação dos informantes desses dados, uma vez que os dados tabulados correspondem a estimativas obtidas a partir da agregação de unidades investigadas na amostra ponderadas pelos fatores de expansão (ou pesos), inerentes ao plano amostral da pesquisa.

Já para a divulgação de microdados, o IBGE disponibiliza arquivos com microdados desidentificados, i.e., sem variáveis que propiciem a identificação imediata do informante, como seu nome ou endereço. O Instituto argumenta outra vez que, dada a natureza da investigação, a simples desidentificação dos registros já seria o suficiente para a garantir a impossibilidade de reidentificação [34]: [7]

> O procedimento de amostragem, por si só, já se configura numa técnica de tratamento estatístico para o controle do risco de revelação de dados individuais.

Censos Demográficos Decenais No que se refere aos Censos Demográficos Decenais, o IBGE divulgada microdados apenas sobre a parcela de investigação por amostragem. Já para os dados censitários dos Censos, o IBGE não publica microdados, mas apenas dados agregados, uma vez que seria possível a reidentificação dos informantes indiretamente, com ou sem o auxílio de informações externas. Para os dados agregados, o IBGE afirma não aplicar procedimentos de desidentificação às células das tabelas. Entretanto, para o Censo Demográfico de 2010, setores com menos do que cinco domicílios particulares permanentes tiveram os valores omitidos para a maioria das variáveis.

[6] Os autores do presente documento acreditam que as afirmações do IBGE quanto ao baixo risco à privacidade acarretado pela divulgação de dados agregados merecem escrutínio cuidadoso. A literatura técnica é rica em exemplos de violações de privacidade decorrentes da divulgação de dados agregados, e esse fato é reconhecido no próprio *Documento de Trabalho sobre Política Estatística 22 - Relatório sobre a Metodologia de Limitação de Divulgação Estatística* das agência federais americanas [93], assim como no relatório sobre o Censo de 2020 nos EUA [57], ambos já tratados no presente documento.

[7] Os autores do presente documento acreditam que também as afirmações do IBGE quanto ao baixo risco à privacidade acarretado pela divulgação de dados amostrais merecem escrutínio cuidadoso, uma vez que a literatura técnica contém exemplos de violações de ocorridas em divulgações desse tipo [76].

Finalmente, o IBGE conta com uma Sala de Acesso a Dados Restritos (SAR), na qual o público externo tem acesso aos microdados de forma controlada, de acordo com o perfil do usuário que acessa o serviço.

3.3 Estudo de caso nacional 2: Divulgação de dados pelo Portal da Transparência do Governo Federal

De acordo com o próprio Governo Federal [48], o Portal da Transparência por ele mantido [8] é um canal pelo qual o cidadão pode acompanhar a utilização dos recursos federais arrecadados com impostos no fornecimento de serviços públicos à população, além de informar-se sobre outros assuntos relacionados à Administração Pública Federal. O site, lançado em 2004 e remodelado em junho de 2018, é mantido pela Controladoria Geral da União (CGU) e tem o objetivo de garantir e ampliar a transparência da gestão pública e fortalecer a participação social na fiscalização dos gastos e investimentos do Poder Executivo Federal.

Regulamentação específica Aplicam-se ao Portal da Transparência do Governo Federal a Lei de Acesso à Informação (LAI) e a Lei Complementar 131/2009, também conhecida como Lei Capiberibe [46], que permite o acesso a informações referentes ao Poder Executivo Federal. Em particular, a Lei Capiberibe estabelece que devem publicizar-se todos os atos praticados pelas unidades gestoras no decorrer da execução de despesa, no momento de sua realização, com a disponibilização mínima dos dados referentes ao número do correspondente processo, ao bem fornecido ou ao serviço prestado, à pessoa física ou jurídica beneficiária do pagamento e, quando for o caso, ao procedimento licitatório realizado.

Dados divulgados Dentre os dados disponibilizados, encontram-se microdados referentes a fichas de remuneração de servidores públicos e recursos disponibilizados e sacados por beneficiários de programas sociais, como o Bolsa Família, o Benefício de Prestação Continuada, e o Auxílio Emergencial devido à pandemia de COVID-19, implementado em abril de 2020. Em ambos os casos, é possível encontrar para um mesmo indivíduo dados como nome completo, CPF parcial, estado e cidade de residência, além dos valores recebidos. Exceto pela omissão parcial do CPF, nenhuma outra técnica de controle de divulgação foi aplicada aos microdados, sendo possível a reidentificação direta dos indivíduos presentes nos registros.

[8] http://www.portaltransparencia.gov.br/

4 Análise da atual situação do Inep frente aos panoramas internacional e nacional e à literatura técnica em privacidade

Nesta seção apresentamos uma análise da atual forma de divulgação dos Censos Educacionais por parte do Inep, com especial foco nos possíveis riscos à privacidade dela decorrentes. Iniciamos a seção com uma contextualização da atual situação do Instituto, que inclui um sumário dos principais pontos da regulamentação específica aplicável ao mesmo, uma breve descrição da atual forma de divulgação das Pesquisas Educacionais realizadas pelo Inep e a identificação das atuais metodologias de mitigação de danos à privacidade aplicadas pelo Instituto, Em seguida, apresentamos um sumário comparativo do contexto do Inep com o dos outros casos de estudo internacionais e nacionais já apresentados neste documento e destacamos os pontos identificados como os mais relevantes à atual situação do Inep. Por fim, realizamos uma avaliação preliminar dos possíveis riscos decorrentes da forma de divulgação adotada pelo Instituto de acordo com o conhecimento acumulado na literatura técnica sobre proteção de privacidade.

4.1 Descrição da atual forma de divulgação dos Censos Educacionais pelo Inep

A harmonização dos requisitos legais da LAI e da LGPD são desafios para o Inep, em particular, na divulgação dos dados de seus Censos Educacionais. [1]

É relevante observar que existem regulamentos institucionais que visam a dar transparência à forma como o Inep produz e divulga os resultados das pesquisas que desenvolve.

[1] A Seção 3.1 apresenta uma visão geral destas leis, enquanto os Anexos B e C deste documento complementam esta exposição com uma seleção de trechos, respectivamente, da LAI e da LGPD identificados como sendo de impacto mais relevante para a divulgação dos Censos Educacionais por parte Inep.

Mais precisamente, a Portaria 91/2018 [64] apresenta um conjunto de princípios fundamentais e de boas práticas que orientam a produção e divulgação das estatísticas educacionais oficiais produzidas pelo Instituto. Já a Portaria 492/2018 [65] institui a *Política de Divulgação de Estatísticas, Exames e Avaliações, Estudos e Pesquisas Educacionais do Inep*. Em particular, em seu Artigo 6º, a Portaria define que "os dados pessoais coletados pelo Inep para fins de pesquisa e produção de estatísticas serão empregados apenas para os fins a que se destinam, tratados com o devido sigilo estatístico, nos termos da legislação aplicável". Na prática, entretanto, as Portarias acima servem principalmente como guias de políticas institucionais que deveriam se desdobrar em protocolos e procedimentos, mas que ainda não tiveram essas regulamentações definidas.

4.1.1 Regulamentação específica aplicável

A Lei 9.448/1997 [45], assim como o Decreto 6.317/2007 [38], definem como finalidades do Inep, dentre outras, o desenvolvimento e a manutenção de sistemas de estatísticas educacionais e de projetos de avaliação educacional, assim como a disseminação dessas informações e de demais produtos relacionados. Entretanto, diferentemente do mandato do Instituto Brasileiro de Geografia e Estatística (IBGE) para coleta de dados, baseado em Lei e regulamentado por Decreto específicos (ver Seção 3.2 para mais detalhes sobre o caso do IBGE), o Inep dispõe apenas do Decreto 6.425/2008 [39], o qual regulamenta o Censo Anual da Educação.

De acordo com o Decreto 6.425/2008, tanto os estabelecimentos de Educação Básica quanto os estabelecimentos de Ensino Superior, sejam eles públicos ou privados, são obrigados a fornecer as informações solicitadas pelo Inep. Além disso, o Decreto assegura o sigilo dos dados pessoais coletados e veda a utilização dos mesmos para outros fins. Entretanto, o Decreto apenas regulamenta a Lei 9.394/1996 [44], a qual estabelece um mandato à União, e não ao Inep, para a coleta, análise, e disseminação de informações sobre a educação, assim como o acesso a todos os dados e informações necessários de todos os estabelecimentos e órgãos educacionais.

4.1.2 Atual forma de divulgação dos Censos Educacionais

Tendo em vista tanto as finalidades conferidas ao Inep quanto os requisitos da LAI, o Instituto publicou em julho de 2016 a Portaria 370, que institui a *Política de Dados Abertos* [31]. De acordo com o Plano Institucional de Dados Abertos estabelecido, dada a Matriz de Responsabilidade para divulgação de produtos na forma de bases de dados, o Inep produz e publica os seguintes produtos formados por microdados, ou seja, dados na menor unidade de agregação possível:

- **Censo da Educação Básica:** com frequência anual, é a principal pesquisa estatística educacional do país, a qual abrange a Educação Básica e Profissional. Inclui a Educação Infantil, os Ensinos Fundamental e Médio, a Educação Especial,

a Educação de Jovens e Adultos, e os Cursos Técnicos e Cursos de Formação Inicial Continuada ou Qualificação Profissional.

- **Censo da Educação Superior:** com frequência anual, é a mais completa pesquisa estatística sobre as Instituições de Ensino Superior (IES's) do país, as quais ofertam Cursos de Graduação e Sequências de Formação Específica. Inclui informações sobre as IES, além de seus alunos e docentes.
- **Exame Nacional do Ensino Médio (ENEM):** com frequência anual, é responsável por avaliar o desempenho escolar ao final da Educação Básica. Seus resultados colaboram para o acesso à Educação Superior, incluindo a oferta de financiamento e apoio estudantil.
- **Sistema de Avaliação da Educação Básica (SAEB):** com frequência bienal, consiste em um conjunto de testes e questionários aplicados na rede pública e em uma amostra da rede privada. Auxilia o Inep na realização de um diagnóstico da Educação Básica e de possíveis fatores que interfiram no desempenho do estudante.
- **Exame Nacional de Certificação de Competências de Jovens e Adultos (ENCCEJA):** com frequência eventual, afere competências, habilidades, e saberes de jovens e adultos que não tenham concluído o Ensino Fundamental ou o Ensino Médio na idade adequada.
- **Exame Nacional de Desempenho dos Estudantes (ENADE):** com frequência anual, avalia o rendimento dos concluintes dos Cursos de Graduação tanto em relação aos conteúdos programáticos previstos nas diretrizes curriculares dos cursos, quanto no que diz respeito às habilidades desenvolvidas e à atualização do aluno em relação ao Brasil e ao mundo.

Os microdados das Pesquisas acima são divulgados pelo Inep para *download* no domínio do Instituto, [2] com links também disponibilizados a partir do Portal Brasileiro de Dados Abertos [3] e em mecanismos de pesquisa como o Google Dataset Search. [4]

Para o Censo da Educação Básica, o Inep disponibiliza dados de 1995 até 2019, assim como para o Censo da Educação Superior, exceto pela base de dados referente a 2019, ainda não publicada. A partir de 2007, o Instituto passou a divulgar microdados relativos ao Censo da Educação Básica com informações individuais sobre alunos e professores, além de escolas e turmas. O mesmo ocorreu a partir de 2009 para o Censo da Educação Superior, com a divulgação de microdados sobre alunos e professores, além de instituições e cursos.

No caso dos exames, o Inep disponibiliza microdados com informações individuais sobre os participantes do ENEM desde 1998 até 2018, e do ENADE desde 2004 até 2018.

[2] http://inep.gov.br/microdados
[3] http://dados.gov.br/
[4] https://datasetsearch.research.google.com/

Vale destacar que as respostas dadas por cada um dos participantes ao Questionário Socieconômico também constam dos microdados, assim como as alternativas por eles selecionadas para cada uma das questões do exame.

Especificamente para o ENEM, vale ainda destacar que o Inep passou a divulgar o Código da Escola de cada participante a partir de 2001, com o cuidado de utilizar uma máscara de proteção e não o Código da Escola conforme publicado no Censo da Educação Básica. Essa medida vigorou até a publicação dos microdados para o ENEM de 2006, sendo que a partir de 2007, o Código da Escola passou a apresentar o valor real conforme o Censo da Educação Básica. Além disso, o Inep publicou uma segunda edição para várias das bases de microdados do ENEM em 2019, sendo a remoção da data de nascimento e a inclusão da idade dos inscritos a alteração mais relevante.

4.1.3 Métodos de proteção de privacidade atualmente empregados

A partir das análises iniciais realizadas sobre os arquivos de microdados dos Censos Educacionais disponibilizados publicamente, identificamos que o Inep aplica as seguintes técnicas de proteção de privacidade aos microdados divulgados:

- *desidentificação*, pela qual são removidos possíveis identificadores individuais óbvios dos registros (como nome, CPF, RG, ou endereços em níveis mais detalhados que as cidades); e
- *pseudonimização*, em que o Instituto atribui a cada registro um código único de identificação artificialmente criado, substituindo identificadores individuais naturais (como nome, CPF ou RG).

Nenhuma outra medida para o controle de divulgações estatísticas foi identificada.

4.2 Sumários comparativos da atual situação do Inep com os estudos de caso internacionais e nacionais

Em seções anteriores deste documento apresentamos uma coleção de estudos de caso relativos a como institutos produtores de estatísticas oficiais equilibram requisitos de transparência e de proteção de privacidade em seus métodos de divulgação de informação.

Em particular, apresentamos tanto casos internacionais, no contexto dos Estados Unidos da América, da União Europeia, e da Austrália (Seção 2), quanto casos nacionais, com foco na divulgação de dados por parte do IBGE e do Portal da Transparência do Governo Federal (Seção 3). Na presente seção provemos sumários comparativos entre a atual situação do Inep em relação à divulgação das informações de seus Censos Educacionais e estes estudos de caso, organizados da seguinte forma:

- a Tabela 4.2.1 apresenta um quadro-sumário dos estudos de caso internacionais no contexto dos EUA;
- a Tabela 4.2.2 apresenta um quadro-sumário dos estudos de caso internacionais no contexto da União Europeia e Austrália; e
- a Tabela 4.2.3 apresenta um quadro-sumário dos estudos de caso nacionais, incluindo a atual situação do Inep em relação à divulgação de seus Censos Educacionais.

4.3 Destaques da comparação de estudos de caso identificados como relevantes à atual situação do Inep

A partir dos quatro critérios utilizados na análise da seção anterior, identificamos os seguintes destaques em relação à atual forma de divulgação dos Censos Educacionais pelo Inep frente àquela adotada nos panoramas internacional e nacional levantados.

Principais pontos da regulamentação geral relevante Todos os países e regiões internacionais pesquisados (EUA, União Europeia e Austrália) possuem alguma forma de regulamentação geral que define parâmetros mínimos de cuidado na divulgação de dados pessoais e estabelece um conjunto de direitos a indivíduos, incluindo aqueles sobre coleta, manutenção e tratamento dos dados, com especial foco na preservação de privacidade de dados pessoais.

Destaque identificado relevante à atual situação do Inep: A LGPD brasileira foi fortemente inspirada pelo seu equivalente europeu, o GDPR, com o qual compartilha filosofia e abrangência similares. De fato, pode-se constatar que a legislação brasileira relativa à privacidade se aproxima mais da europeia, que por sua vez é muito diferente da legislação dos EUA. Entretanto, há significativas diferenças entre a LGPD e a GDPR, e o Anexo D apresenta um paralelo mais detalhado dos dois regulamentos.

Principais pontos da regulamentação específica relevante aplicável à agência ou órgão Todos os casos internacionais pesquisados exigem a utilização dos dados para o propósito original para que foram coletados, com raras exceções em caso de segurança nacional (como no caso de uso de dados para combate ao terrorismo nacional ou internacional, permitido pelo *USA PATRIOT Act* dos EUA). Além disso, a maioria dos casos internacionais pesquisados prevê a obrigatoriedade do uso de pelo menos técnicas simples de preservação de privacidade de dados individuais, como alguma forma de ocultação de identificadores óbvios via *desidentificação*, ou substituição de identificadores naturais por códigos artificialmente gerados via *pseudonimização*.

Destaque identificado relevante à atual situação do Inep: Não foi identificada regulamentação específica que obrigue o Inep a manter a divulgação de todos os microdados de seus Censos Educacionais da forma como atualmente é feita. A manutenção da acurácia dos dados colhidos pelo Inep depende da confiança que as fontes dos mesmos depositam nos processos empregados pelo Instituto. É crucial, portanto, zelar pela percepção dos titulares dos dados –incluindo escolas professores, estudantes e seus familiares– quanto à apropriada proteção de sua privacidade.

Formas de divulgação de dados No caso das escolas e do Ministério da Educação da Holanda, não foi identificada a divulgação de microdados sobre indivíduos. No contexto da Austrália ocorreu, até o último Censo realizado, em 2016, a publicação de microdados apenas de forma controlada. No contexto dos EUA, em geral, agências federais devem dar preferência a dados agregados, mas até recentemente microdados ainda eram publicados após submetidos a medidas de proteção como: (i) divulgação de dados de apenas uma amostra da população; (ii) aplicação de técnicas de proteção como introdução de ruído e permutação; ou (iii) autorização de acesso completo apenas através de salas seguras.

No contexto nacional, o IBGE publica tanto microdados quanto dados agregados para as Pesquisas Domiciliares por Amostragem Probabilística, enquanto que para os Censos Demográficos Decenais são divulgados microdados apenas sobre a parcela de investigação por amostragem, sendo os dados censitários divulgados apenas de forma agregada. Já o Portal da Transparência do Governo Federal, por seu próprio objetivo expresso, publica microdados de forma explícita, incluindo dados pessoais. É relevante ressaltar que essa divulgação por parte do Portal da Transparência pode ser utilizada como informação auxiliar em ataques de reidentificação a outras bases de dados tratadas com as técnicas de desidentificação e, em especial, de pseudonimização. (Na Seção 4.4 exemplificamos casos de ataques que se valem do cruzamento de informação entre bases distintas.) É, portanto, de particular importância que as entidades que divulgam microdados de pesquisas, incluindo o Inep, considerem esse risco. No contexto do Inep a preocupação é reforçada pela grande abrangência da população brasileira cujos dados estão contidos no Portal da Transparência (e.g., sabe-se que parte considerável dos alunos das escolas públicas brasileiras –algo entre 40% e 50%– consiste em beneficiários do Bolsa Família e, portanto, seus dados estão disponíveis no Portal da Transparência).

Destaque identificado relevante à atual situação do Inep: Frente a todos os casos internacionais estudados, o Inep publica em seus Censos Educacionais a maior quantidade de microdados individuais e em maior nível de detalhes.

Medidas de proteção de privacidade empregadas No caso das publicações dos institutos educacionais e do Ministério da Educação da Holanda, não foi identificada a

presença de microdados, sob nenhuma forma, o que indica a forte preocupação destes órgãos com os riscos de violação de privacidade inerentes a esse tipo de dado. O Escritório Australiano de Estatística (ABS) protege a publicação de microdados com divulgação de apenas amostras da população, acesso restrito aos microdados baseado em pré-aprovação, e cobrança monetária por este acesso.

Nos EUA, as agências federais produtoras de estatísticas oficiais reconhecem explicitamente, pelo menos desde 2005, a dificuldade inerente de se protegerem microdados contra violações de privacidade, uma vez que há sempre a possibilidade de que informações externas aos microdados sejam utilizadas para a reidentificação de indivíduos. Por esta razão, a divulgação de microdados ocorria de forma controlada por medidas que têm como potencial efeito colateral uma significativa diminuição da precisão dos dados divulgados. (Por exemplo, os registros submetidos a permutação, a apagamento e imputação, ou a *blurring* não são marcados de qualquer forma nos arquivos divulgados, de modo a garantir a segurança desses registros; com isso, um elevado grau de incerteza é adicionado aos dados, uma vez que o usuário se torna incapaz de determinar se um registro isolado contém dados reais, permutados, imputados, ou ofuscados.) Em particular, o Escritório do Censo dos EUA (USCB) reconheceu recentemente que mesmo as medidas adotadas até o último Censo Decenal, em 2010, para proteção de privacidade não são suficientes. O USBC está no momento alterando drasticamente a forma de divulgação de dados de seu Censo de 2020 para dados tabulares protegidos por privacidade diferencial. Além disso, usam-se nos EUA salas seguras (*Research Data Centers* - RDCs) espalhadas pelo país para acesso controlado a microdados por indivíduos previamente autorizados.

No contexto nacional, o IBGE reconhece os riscos inerentes à divulgação de microdados, publicando aqueles relativos às Pesquisas Domiciliares por Amostragem Probabilística apenas após desidentificação. Entretanto, o IBGE é mais permissivo na publicação de dados agregados, afirmando não aplicar medidas extras de proteção neste caso sob o argumento de que a reidentificação de informantes não seria possível dada a natureza da investigação e da divulgação agregada. [5] Já o Portal da Transparência do Governo Federal, por seu próprio objetivo expresso, publica microdados de forma explícita, sem aparente preocupação com a consequente reidentificação de indivíduos possível.

Destaque identificado relevante à atual situação do Inep: Em todos os casos internacionais estudados, os órgãos produtores de estatísticas oficiais adotam medidas mais rígidas que o Inep no controle de divulgação de microdados, com destaque para: (i) não publicação de microdados individuais de estudantes sob nenhuma forma (caso

[5] Como já mencionado na Seção 3.2, os autores do presente documento acreditam que as afirmações do IBGE quanto ao baixo risco à privacidade acarretado pela divulgação de dados agregados merecem escrutínio cuidadoso. A literatura técnica é rica em exemplos de violações de privacidade decorrentes da divulgação de dados agregados, e esse fato é reconhecido no próprio *Documento de Trabalho sobre Política Estatística 22 - Relatório sobre a Metodologia de Limitação de Divulgação Estatística* das agência federais americanas [93], assim como no relatório sobre o Censo de 2020 nos EUA [57], ambos já tratados no presente documento.

das escolas e do Ministério da Educação na Holanda); (ii) publicação de microdados apenas por amostras da população (caso do Escritório do Censo dos EUA - USCB e do Escritório Australiano de Estatística - ABS); ou (iii) acesso a todo o conjunto de microdados permitido apenas sob autorização prévia e em salas seguras (também caso do Escritório do Censo dos EUA - USCB). Dado o compromisso não trivial que existe entre transparência em divulgação de dados e proteção de privacidade, as medidas adotadas pelos institutos internacionais estudados em geral diminuem –e, em alguns casos, consideravelmente– a acessibilidade e a qualidade das informações prestadas de forma aberta à sociedade. No caso do Inep, identificamos que a extensa base de microdados dos Censos Educacionais é divulgada em sua totalidade, sendo que os microdados são tratados apenas com as técnicas de *desidentificação*, em que se removem possíveis identificadores individuais óbvios dos registros (como nome, CPF, RG) e de *pseudonimização*, em que tais identificadores individuais óbvios são substituídos por um código único de identificação artificialmente criado. A atual forma de divulgação de microdados pelo Inep está sujeita a vários riscos já identificados na literatura técnica na área de privacidade (que são melhor discutidos na Seção 4.4).

4.4 Possíveis riscos à privacidade que a literatura técnica aponta na atual situação do Inep

O plano de trabalho do presente TED 8750 prevê produtos cujo objeto são análises detalhadas sobre os riscos à privacidade decorrentes da atual forma de divulgação dos dados dos Censos Educacionais pelo Inep (Produto 02), assim como possíveis formas de mitigação dos mesmos (Produtos 03). Nesta seção procedemos a uma análise preliminar da atual situação do Inep baseada em uma revisão dos riscos já documentados na literatura.

A literatura técnica na área de proteção de privacidade apresenta fartos exemplos de que métodos de mitigação de riscos à privacidade como aqueles atualmente empregados pelo Inep –mais especificamente, o método de *desidentificação* (i.e., remoção de identificadores individuais óbvios, como CPF, nome completo, ou identidade) e o método de *pseudonimização* (i.e., substituição dos identificadores usuais por identificadores artificiais, como números únicos produzidos para cada indivíduo)– são, em isolamento, insuficientes para proteger a privacidade dos titulares dos dados. Podemos destacar os seguintes trabalhos como alguns dos mais influentes sobre o assunto nas últimas décadas:

(I) Em 1998, Samarati e Sweeney observaram que, mesmo quando os titulares de dados removem ou criptografam identificadores explícitos (como nomes, endereços e números de telefone) de indivíduos aos quais microdados divulgados se referem, outros dados distintos, que elas denominaram *quaseidentificadores* (*quasi-identifiers*, no original em inglês), geralmente se combinam de maneira inadequada e podem ser vinculados a informações publicamente disponíveis para reidentificar os indivíduos

[105]. Como exemplo de quaseidentificadores poderosos, Sweeney demonstrou em 2000 que 87% da população dos EUA poderia ser unicamente identificada usando-se apenas o *zip-code* (o equivalente americano ao CEP brasileiro), o gênero e data de nascimento do indivíduo [109]. Portanto, qualquer divulgação de microdados que contivesse pelo menos estes três campos (e.g., algum censo ou base de dados médicos) permitiria a reidentificação de até 87% da população americana, levando a uma consequente exposição generalizada de atributos sensíveis individuais.

(II) Em 2008, Narayanan e Shmatikov apresentaram uma nova técnica de reidentificação estatística de microdados de alta dimensão (e.g., aqueles que contêm preferências individuais, recomendações, ou registros de transações) e a aplicaram ao conjunto de microdados do *Prêmio Netflix* [77], que continha as notas atribuídas aos filmes assistidos (*movie ratings*, do original em inglês) por 500.000 assinantes da Netflix, então o maior serviço de aluguel de filmes *on-line* do mundo [76]. Os pesquisadores demonstraram que alguém que conhecesse apenas poucos dados sobre um assinante individual poderia facilmente reidentificar o registro desse assinante no conjunto de microdados. Usando o *Internet Movie Database* (IMDB) [6] como fonte de informação auxiliar, os pesquisadores foram capazes de reidentificar com sucesso nos registros da Netflix usuários famosos, descobrindo suas aparentes preferências políticas e outras informações potencialmente sensíveis.

Como exemplo concreto dos riscos apresentados pelos trabalhos acadêmicos acima, destacamos o caso de reidentificação da AOL [2]. Esse caso ocorreu em um contexto bastante similar ao do Inep no momento, e serve como ilustração dos riscos à privacidade decorrentes do uso de apenas métodos de desidentificação e pseudonimização. Em 4 de agosto de 2006, a *AOL Research* disponibilizou um arquivo de texto compactado em um de seus sites contendo cerca de 20 milhões de palavras-chave usadas por 650.000 usuários em seu motor de busca, em um período de três meses, destinado a fins de pesquisa. O arquivo foi desidentificado e pseudonimizado de forma que identificadores pessoais óbvios foram removidos e nomes foram substituídos por números como pseudônimos, mas as palavras-chave das consultas realizadas por cada indivíduo permaneceram no arquivo. Como exemplo, os dados divulgados mostravam que o indivíduo #4.417.749 havia realizado pesquisas no motor de busca por expressões como *"dedos dormentes"*, *"homens solteiros com mais de 60 anos"*, *"cão que urina em tudo"*, e *"paisagistas em Lilburn, GA"*. Cruzando os dados dessas pesquisas com informações de listas telefônicas, o jornal *The New York Times* foi capaz de identificar unicamente o indivíduo #4.417.749 como Thelma Arnold, uma viúva de 62 anos de Lilburn, estado da Georgia, EUA [106]. Apesar de, no caso específico de Arnold, o jornal ter divulgado sua identidade com sua permissão, o risco potencial às outras centenas de milhares de usuários da base estava escancarado. Essa violação foi amplamente divulgada na imprensa, e levou à renúncia da então CTO da organização, Maureen Govern, em 21 de agosto de 2006. A AOL reconheceu que a liberação dos dados, mesmo desidentificados e pseudonimizados, foi um erro e os removeu

[6] https://www.imdb.com/

de sua página; no entanto, a remoção foi tardia demais. Os dados foram redistribuídos por outras pessoas e ainda podem ser baixados de páginas-espelho.

É importante observar que a atual forma de divulgação de microdados dos Censos Educacionais por parte do Inep é muito similar àquela adotada pela AOL no exemplo acima, o que suscita a razoável preocupação sobre em qual extensão os dados do Inep estariam sujeitos a violações de privacidade semelhantes.

De fato, como mencionamos na introdução do presente documento, os pesquisadores brasileiros Queiroz e Motta já observaram a vulnerabilidade da atual situação do Inep em um estudo de 2015 [102]. Utilizando o arquivo referente aos docentes do Censo da Educação Superior do ano de 2013, eles conseguiram reidentificar unicamente um dos autores do estudo dentre todos os 383.683 registros de docentes na base utilizando apenas sua data de nascimento, gênero, e nome da Instituição de Ensino Superior à qual estava vinculado. Como possível solução, Queiroz e Motta sugeriram a aplicação de dois métodos mais sofisticados aos Censos Educacionais do Inep: os métodos sintáticos conhecidos como *k-anonymity* [105] e *l-diversity* [71], o que foi feito com o auxílio da ferramenta ARX [100]. Os resultados preliminares apresentados pelos autores poderiam parecer, à primeira vista, promissores: utilizando a ferramenta ARX, eles computaram que o número de indivíduos com alto risco de reidentificação caiu de 100% no conjunto de microdados original para 50% no conjunto de microdados tratado com as novas técnicas. Entretanto, métodos sintáticos como k-anonymity e l-diversity são demonstradamente vulneráveis a um tipo de ataque conhecido como *ataque composicional*, ao qual o Inep estaria sujeito por publicar os Censos Educacionais com frequência regular. A fragilidade destes métodos já foi demonstrada pelo próprio ataque de Narayanan e Shmatikov citado no Exemplo II acima, relativo ao Prêmio Netflix (que foi realizado sobre uma base k-anonimizada), e confirmada pela avaliação da equipe responsável pelo novo método de publicação do Censo 2020 dos EUA [57]. Portanto, mesmo os métodos sugeridos por Queiroz e Motta em seu trabalho de 2015, isoladamente, não resolveriam de forma definitiva o problema de proteção de privacidade no caso do Inep.

Ademais, conforme mencionado na Seção 4.3, os microdados pessoais divulgados pelo Portal da Transparência poderiam ser usados como informação auxiliar em tentativas de reidentificação de indivíduos nos Censos Educacionais divulgados pelo Inep. Mais precisamente, os dados do Portal poderiam ser cruzados com os Censos de forma análoga a como os dados do IMDB foram cruzados com os dos usuários da Netflix no caso descrito acima. No contexto do Inep a preocupação é reforçada pela grande abrangência da população brasileira cujos dados estão contidos no Portal da Transparência (e.g., parte considerável dos alunos das escolas públicas brasileiras consiste em beneficiários do Bolsa Família e, portanto, seus dados estão disponíveis no Portal da Transparência).

A discussão da presente seção demonstra não apenas que há uma possibilidade concreta de violações de privacidade decorrentes da atual forma de divulgação de dados dos Censos Educacionais por parte do Inep, como também o fato de que a escolha das técnicas de mitigação de danos à privacidade devem ser adotadas com critérios cuidadosos.

	Estudos de caso internacionais: Estados Unidos da América		
	Caso 1: Agências Federais (2005) - Seção 2.2.2	**Caso 2: Centro Nacional de Estatísticas da Educação (NCES) - Seção 2.2.3**	**Caso 3: Escritório do Censo (USCB) - Seção 2.2.4**
Principais pontos da regulamentação geral relevante	**Lei de Privacidade de 1974, Tít. 5 Cód. EUA** [89]: agências federais que lidam com registros de informações pessoais identificáveis devem protegê-los contra uso indevido, e seu uso para pesquisas estatísticas deve garantir que os mesmos não sejam individualmente identificáveis. **Lei de Proteção à Informação Confidencial e Eficiência Estatística de 2002 [92]**: torna-se crime federal a divulgação intencional, sem consentimento, de dados individualmente identificáveis obtidos para fins estatísticos sob promessa de confidencialidade.		
Principais pontos da regulamentação específica relevante aplicável à agência ou órgão	**Documento de Trabalho sobre Política Estatística 22 - Relatório sobre a Metodologia de Limitação de Divulgação Estatística de 1994, revisado em 2005 [93]**: define diversos métodos de limitação de divulgação estatística, como também as práticas adotadas por 14 agências federais dos EUA à época. Além disso, reconhece a dificuldade inerente à proteção de microdados contra divulgação indevida, uma vez que há sempre a possibilidade de que informações externas a tais dados sejam utilizadas para a reidentificação de indivíduos.	**Documento 22** [93]: aplica-se ao NCES como agência federal. **Lei de Reforma das Ciências da Educação de 2002** [87]: informações individualmente identificáveis sobre estudantes, suas famílias e escolas, devem permanecer confidenciais. ***USA PATRIOT Act* de 2001** [91]: o Procurador Geral pode ter acesso judicial a informações individualmente identificáveis caso relacionadas a terrorismo. **Padrões Estatísticos do Centro Nacional de Estatísticas da Educação de 2012** [94]: todos os microdados de uso público devem ser tratados para limitar a possibilidade de reidentificação auxiliada por informações externas.	**Documento 22** [93]: aplica-se ao USCB como agência federal. **Código dos EUA, Título 13** [88]: toda informação deve ser utilizada exclusivamente para os propósitos estatísticos para os quais foi coletada, e nenhum indivíduo ou estabelecimento pode ter sua identidade revelada. **Padrões de Qualidade Estatística de 2013** [22]: técnicas de prevenção devem ser utilizadas para impedir a liberação de informações protegidas, particularmente aquelas de pessoas ou estabelecimentos.
Formas de divulgação de dados	**Microdados**: em nível de registros individuais (mas devem ser evitados). **Dados agregados de frequência**: relacionados a contagens (e.g., número de lojas em uma região). **Dados agregados de magnitude**: relacionados a grandezas como lucro (e.g., número de lojas e a receita bruta agregada).	O NCES disponibiliza **microdados de acesso restrito** para agentes autorizados e publica **microdados de uso público** produzidos a partir dos primeiros, assim como ferramentas de análise *on-line*, nas quais os microdados são agregados para a produção de estimativas tabulares ou, em alguns casos, para análises de regressão.	**Censos 1960–2010**: dados **agregados de magnitude e frequência**, e **microdados** referentes a uma **pequena amostra da população** (*Public Use Microdata Samples* - PUMS), sob diferentes formas de proteção. **Censo 2020: agregação tabular** usando **privacidade diferencial (*differential privacy*)**.
Medidas de proteção de privacidade empregadas	**Aplicadas a microdados:** (i) divulgação por amostras; (ii) remoção de identificadores óbvios; (iii) limitação do detalhamento geográfico; e (iv) limitação de quantidade/detalhamento de valores de variáveis. Medidas adicionais em variáveis de alto risco: (v) *top-* e *bottom-coding*; (vi) recodificação em intervalos ou arredondamentos; (vii) adição de ruído; (viii) permutação de valores; (ix) troca de classificação; (x) apagamento e imputação; (xi) *blurring*; (xii) supressão direcionada. **Aplicadas a dados agregados de magnitude e frequência**: (i) (n,k)-*rule*; (ii) *p-percent rule*; (iii) *threshold Rule (N+)*; (iv) truncamento; (v) permutação; (vi) recodificação; (vii) adição de ruído; e (viii) controle de consultas.	**Aplicadas a microdados de uso público acessados por ferramentas de análise *on-line* providas pelo NCES**: perturbação via (i) *threshold rule (N+)*; (ii) truncamento, (iii) permutação; e (iv) recodificação. **Aplicadas a microdados de acesso restrito acessíveis pelas mesmas ferramentas**: (v) o tamanho exato da amostra não pode ser divulgado; (vi) as edições de confidencialidade devem ter sido aprovadas pelo DRB (*Disclosure Review Board*) responsável; (vii) a ferramenta *on-line* deve divulgar apenas contagens ponderadas.	**Aplicadas a microdados dos Censos 1960–2010**: (i) truncamento e recodificação; (ii) apagamento e imputação; (iii) permutação; (iv) truncamento de variáveis categóricas; (v) adição de ruído; (vi) uso de dados parcialmente sintéticos. **Aplicadas ao Censo 2020**: substituição de métodos sintáticos por privacidade diferencial (*differential privacy*) [51], via o algoritmo *TopDown* de geração de tabelas [8]. (Entretanto, o novo método enfrenta dificuldades técnicas [58] e de comunicação com a comunidade de estatística [73].) **Salas seguras (*Research Data Centers* - RDC's)**: salas espalhadas pelo país para acesso a microdados por indivíduos autorizados, de forma controlada.

Tabela 4.2.1: Sumário comparativo dos estudos de caso internacionais efetuados no contexto dos EUA

	Estudos de caso internacionais: União Europeia e Austrália	
	Caso 4: Divulgação de dados educacionais por escolas e pelo Ministério da Educação da Holanda - Seção 2.3.2	**Caso 5: Escritório Australiano de Estatística (ABS) - Seção 2.4.2**
Principais pontos da regulamentação geral relevante	**Regulamento UE 2016/679 (*General Data Protection Regulation* - GDPR)** [29]: unifica o quadro regulamentar europeu e provê aos cidadãos e residentes da UE formas de controlar seus dados pessoais, conferindo-lhes os seguintes direitos básicos [99]: (i) de informação; (ii) de acesso; (iii) de retificação; (iv) de apagamento dos dados; (v) à limitação de tratamento; (vi) de portabilidade; (vii) a não ficar sujeito a decisões individuais automatizadas; e (viii) de oposição.	**Princípios Australianos de Privacidade, Lei de Privacidade de 1988, revisada em 2015** [81]: garantia do uso de *anonimização* (i.e., a desassociação dos indivíduos de seus respectivos registros) ou de *pseudonimização* (i.e., a atribuição de um código individual a cada registro de uma base de dados), além da transparência por parte das entidades governamentais na coleta e no tratamento de informações pessoais, assim como de regras para o uso e a divulgação de dados pessoais.
Principais pontos da regulamentação específica relevante aplicável à agência ou órgão	**Lei *Algemene Verordening Gegevensbescherming* (AVG) de 2018** [4]: implementa a GDPR europeia. **Autoridade de Privacidade - AP (*Autoriteit Persoonsgegevens*)** [13]: controla o uso e divulgação de informação escolar por institutos de ensino e pelo Ministério da Educação. **Acordo sobre privacidade de 2018** [101]: define que um número único deve ser usado para cada aluno de ensino básico e secundário, servindo como pseudônimo (o que consiste em uma forma de *pseudonimização*). Além disso, estabelece um padrão de protocolos de Internet, [52] e de como as partes devem trocar as informações entre si [69], incluindo uma implementação de referência [68].	**Política de Privacidade do Escritório de Estatística** [82, 86]: a ABS nunca publicou, nem publicará, informações identificáveis sobre indivíduos, residências, ou estabelecimentos comerciais.
Formas de divulgação de dados	**Dados divulgados por escolas**: microdados dos alunos nunca são publicados, em nenhuma forma. **Dados divulgados pelo Ministério da educação**: microdados dos alunos nunca são publicados, em nenhuma forma; são divulgados, de forma *on-line*, apenas dados de escolas ou regiões, e dados de alunos apenas em forma agregada.	**Dados relavitos ao censo**: Microdados de apenas amostras da população do censo estão disponíveis. **Dados relativos ao censo longitudinal**: Base com microdados do censo longitudinal (*Australian Census Longitudinal Dataset* - ACLD) é criada a partir de amostras aleatórias com correspondência e inclui: (i) 2006–2011 ACLD: amostra de 5% de registros aleatórios, o equivalente a aproximadamente um milhão de registros, do Censo de 2006 com os registros correspondentes do Censo de 2011; e (ii) 2011–2016 ACLD: aproximadamente 1,2 milhão de registros do Censo de 2011 com os registros correspondentes do Censo de 2016.
Medidas de proteção de privacidade empregadas	**Microdados referentes a alunos nunca são publicados, em nenhuma forma**, nem por escolas nem pelo Ministério da Educação. **Medidas tomadas por instituições de ensino:** mesmo agentes autorizados só têm acesso aos dados realmente necessários para uma dada tarefa, e mesmo assim de forma temporária, controlada e justificada. **Medidas tomadas pelo Ministério da Educação:** são divulgados, de forma *on-line*, apenas dados de escolas ou regiões, e dados de alunos apenas são divulgados em forma agregada (e.g., valores médios).	Até o Censo 2016, o ABS disponibilizou **microdados** apenas para **amostras** e sob **restrições de acesso** [84], incluindo cobrança monetária [85]. Dados do ACLD são acessíveis para usuários registrados no serviço de geração de tabelas e na forma de **microdados para usuários aprovados** que estejam conduzindo projetos consentidos, com cobrança monetária. As únicas medidas de proteção identificadas são a remoção de identificadores óbvios (**desidentificação**) e a **publicação de dados por amostras**.

Tabela 4.2.2: Sumário comparativo dos estudos de caso internacionais efetuados no contexto da União Europeia e Austrália

	Estudos de caso nacionais		
	Caso 1: IBGE Seção 3.2	**Caso 2: Portal da Transparência do Governo Federal Seção 3.3**	**Caso dos Censos Educacionais do Inep - Seção 4**
Principais pontos da regulamentação geral relevante	**Constituição Brasileira de 1988, Art. 5º** [37]: direito à privacidade do indivíduo e direito individual e coletivo à transparência por parte do Estado (mas sem definição de como equilibrar tais princípios entre si). **Lei 12.527/2011 - Lei de Acesso à Informação (LAI)** [47]: o poder público deve garantir o amplo acesso à informação de interesse coletivo ou geral, a qual deve ser disponibilizada via Internet independentemente de requerimento; garante o direito de acesso às demais informações não imediatamente disponibilizadas via Internet; garante o acesso à informação primária, íntegra, autêntica, e atualizada. **Lei 13.709/2018 - Lei Geral de Proteção de Dados Pessoais (LGPD)** [41]: dados anonimizados não são considerados dados pessoais, exceto quando a anonimização puder ser revertida com esforços razoáveis (i.e., considerando fatores objetivos como o custo e o tempo necessários para reverter a anonimização dadas as tecnologias disponíveis e desconsiderando-se o uso de meios de terceiros).		
Principais pontos da regulamentação específica relevante aplicável à agência ou órgão	**Lei 5.534/1968 [43] e Decreto 73.177/1973 [40]**: toda pessoa natural ou jurídica, de direito público ou privado, é obrigada a prestar as informações solicitadas pelo IBGE, mas tais informações são protegidas por sigilo e só podem ser usadas para fins estatísticos [34]. **Código de Boas Práticas das Estatísticas do IBGE de 2013** [33]: (i) o acesso aos microdados não desidentificados deve seguir protocolos de confidencialidade para usuários externos que têm acesso para análise e pesquisa estatística; (ii) o arquivamento das informações pelo Instituto deve seguir protocolos de segurança e confidencialidade estabelecidos.	**Lei Complementar 131/2009 - Lei Capiberibe** [46]: devem-se publicizar todos os atos praticados pelas unidades gestoras no decorrer da execução de despesa, no momento de sua realização, com a disponibilização mínima dos dados referentes ao número do correspondente processo, ao bem fornecido ou ao serviço prestado, à pessoa física ou jurídica beneficiária do pagamento e, quando for o caso, ao procedimento licitatório realizado.	**Lei 9.448/1997** [45] e **Decreto 6.317/2007** [38]: estão entre as responsabilidades do Inep sistemas de estatísticas e projetos de avaliação educacionais, assim como a disseminação dessas informações e de demais produtos relacionados. **Decreto 6.425/2008** [39]: regulamenta o Censo Anual da Educação; estabelecimentos públicos e privados de Educação Básica e Superior devem fornecer as informações solicitadas pelo Inep, sendo garantido o sigilo dos dados pessoais e vedada sua utilização para outros fins. (Entretanto, o Decreto regulamenta a **Lei 9.394/1996** [44], que estabelece à União, e não ao Inep, um mandato para coleta, análise, e disseminação de informações sobre a educação.)
Formas de divulgação de dados	**Guia de procedimentos de 2018** [34]: define os procedimentos adotados pelo IBGE. Em **Pesquisas Domiciliares por Amostragem Probabilística**: são divulgadas tanto informações agregadas em forma de tabelas quanto microdados. Nos **Censos Demográficos Decenais**: publicam-se dados censitários de forma agregada, além de microdados da parcela de investigação por amostragem.	Os dados disponibilizados incluem **microdados** (e.g., referentes a fichas de remuneração de servidores públicos e recursos disponibilizados e sacados por beneficiários de programas sociais, como o Bolsa Família, o Benefício de Prestação Continuada, e o Auxílio Emergencial devido à epidemia de COVID-19 implementado em Abril de 2020).	Divulgam-se **microdados** de Censos, incluindo o da Educação Básica (desde 2008, com dados individuais de alunos, professores, escolas e turmas), o da Educação Superior (desde 2009, com dados individuais de alunos, professores, instituições e cursos), e o ENEM (desde 1998, com dados individuais de participantes, incluindo suas respostas a cada questão do Exame e do Questionário Socieconômico).
Medidas de proteção de privacidade empregadas	**Aplicadas a Pesquisas Domiciliares**: identificadores individuais óbvios são removidos de microdados; dados agregados são reportados sem medidas adicionais. **Aplicadas ao Censos Demográficos Decenais**: IBGE afirma não usar desidentificação, mas no Censo de 2010 setores com menos de cinco domicílios particulares tiveram valores omitidos para muitas variáveis. **Sala de Acesso a Dados Restritos (SAR)**: local onde o público externo tem acesso aos microdados de forma controlada.	Exceto pela omissão parcial do CPF, **nenhuma outra técnica de controle de divulgação foi identificada**, sendo possível uma fácil e imediata reidentificação direta de todos os indivíduos cujos microdados estão presentes nos registros.	Os **microdados** são tratados com as técnicas de **desidentificação**, em que se removem possíveis identificadores individuais óbvios dos registros (como nome, CPF, RG) e de **pseudonimização**, em que tais identificadores individuais óbvios são substituídos por um código único de identificação artificialmente criado.

Tabela 4.2.3: Sumário comparativo dos estudos de caso efetuados no contexto nacional, incluindo o do Inep em sua divulgação dos Censos Educacionais

5 Considerações finais

Neste relatório apresentamos, através de uma coleção de estudos de caso, um panorama internacional e um nacional relativos a como institutos produtores de estatísticas oficiais equilibram requisitos de transparência e de proteção de privacidade em seus métodos de divulgação de informação. Procedemos, então, a uma contextualização da atual situação da divulgação dos Censos Educacionais por parte do Inep frente a esses casos.

Averiguamos que a extensa coleção de microdados dos Censos Educacionais é divulgada em sua totalidade, sendo a ela aplicadas apenas as técnicas de *desidentificação*, em que se removem possíveis identificadores individuais óbvios dos registros (como nome, CPF, RG) e de *pseudonimização*, em que tais identificadores individuais óbvios são substituídos por um código único de identificação artificialmente criado. A análise preliminar desta forma divulgação de microdados revelou que a mesma está sujeita a vários riscos já apontados na literatura técnica na área de privacidade. Dentre estes riscos, destacamos o de reidentificação de indivíduos via cruzamento com informações auxiliares, e a inferência de seus atributos sensíveis. Em nossa visão, tal situação poderia constituir em uma violação da Lei Geral de Proteção de Dados Pessoais.

Em comparação a todos os casos internacionais estudados, averiguamos que o Inep publica em seus Censos Educacionais a maior quantidade de microdados individuais e em maior nível de detalhes. Além disso, em todos os casos internacionais avaliados, os órgãos produtores de estatísticas oficiais adotam medidas mais rígidas do que as do Inep na mitigação de danos à privacidade causados por publicação de microdados. Dentre estas medidas, podemos destacar as seguintes:

(i) não publicação de microdados individuais de estudantes sob nenhuma forma (no caso dos institutos de educação e Ministério da Educação na Holanda);

(ii) publicação de microdados apenas por amostras da população (no caso do Escritório do Censo dos EUA e no do Escritório Australiano de Estatística); ou

(iii) acesso a todo o conjunto de microdados permitido apenas sob autorização prévia, de acordo com os objetivos do pesquisador ou indivíduo interessado, e em salas seguras (também no caso do Escritório do Censo dos EUA).

É amplamente aceito na literatura técnica que existe um compromisso não trivial entre a transparência na divulgação de informação e a proteção de privacidade dos indivíduos envolvidos. Uma consequência desse compromisso, como reconhecido pelo próprio Escritório do Censo dos EUA, é que medidas de mitigação de danos em geral diminuem –e, por vezes, consideravelmente– a qualidade e a utilidade das informações prestadas de forma aberta à sociedade. Acertar esse equilíbrio é um processo árduo e contínuo.

A discussão apresentada neste documento evidencia que não apenas existem significativos riscos de violação de privacidade decorrentes da atual forma de divulgação dos Censos Educacionais pelo Inep, como também que a escolha dos métodos de controle desses riscos deve ser realizada com prudência. Mais precisamente, deve estudar-se sob a luz de critérios técnicos cuidadosos a situação dos Censos Educacionais do Inep em suas peculiaridades, a fim de selecionarem-se técnicas de mitigação de riscos que permitam um compromisso adequado entre a qualidade da informação divulgada e a proteção à privacidade dos indivíduos. Este é exatamente o objetivo geral do presente projeto.

Referências Bibliográficas

[1] Aanslag op het Amsterdams bevolkingsregister. In *Wikipedia: de vrije encyclopedie*. Wikimedia, 2020. Disponível em: https://nl.wikipedia.org/wiki/Aanslag_op_het_Amsterdams_bevolkingsregister.

[2] AOL search data leak. In *Wikipedia: the free encyclopedia*. Wikimedia, 2020. Disponível em: https://en.wikipedia.org/wiki/AOL_search_data_leak.

[3] Privacidade. In *Wikipédia: a enciclopédia livre*. Wikimedia, 2020. Disponível em: https://pt.wikipedia.org/wiki/Privacidade.

[4] Uitvoeringswet Algemene verordening gegevensbescherming. In *Wikipedia: de vrije encyclopedie*. Wikimedia, 2020. Disponível em: https://nl.wikipedia.org/wiki/Uitvoeringswet_Algemene_verordening_gegevensbescherming.

[5] Volkstelling. In *Wikipedia: de vrije encyclopedie*. Wikimedia, 2020. Disponível em: https://nl.wikipedia.org/wiki/Volkstelling.

[6] Wet bescherming persoonsgegevens (Nederland). In *Wikipedia: de vrije encyclopedie*. Wikimedia, 2020. Disponível em: https://nl.wikipedia.org/wiki/Wet_bescherming_persoonsgegevens_(Nederland).

[7] Wet persoonsregistraties. In *Wikipedia: de vrije encyclopedie*. Wikimedia, 2020. Disponível em: https://nl.wikipedia.org/wiki/Wet_persoonsregistraties.

[8] John Abowd, Daniel Kifer, Brett Moran, Robert Ashmead, Philip Leclerc, William Sexton, Simson Garfinkel, and Ashwin Machanavajjhala. Census TopDown: Differentially Private Data, Incrementalschemas, and Consistency with Public Knowledge, 2019. Disponível em: https://columbia.github.io/private-systems-class/papers/Abowd2019Census.pdf.

[9] Agência Brasil. Supremo suspende MP do compartilhamento de dados com IBGE. *UOL*, 2020. Disponível em: https://www.uol.com.br/tilt/noticias/redacao/2020/04/24/supremo-suspende-mp-do-compartilhamento-de-dados-com-ibge.htm.

[10] Australian Bureau of Statistics. Privacy Impact Assessment: Proposal to Retain Name and Address Information from Responses to the 2016 Census of Population and Housing, 2015. Disponível em: https://www.abs.gov.au/websitedbs/D3310114.nsf/4a256353001af3ed4b2562bb00121564/170fd5a4b684aa3eca257f1e0021a392/$FILE/ABS%20Privacy%20Impact%20Assessment%202016%20Census.pdf.

[11] Australian Bureau of Statistics. Retention of names and addresses collected in the 2016 Census of Population and Housing, 2016. Disponível em: https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Retention+of+names+and+addresses+collected.

[12] Australian Privacy Foundation. ABS goes rogue: Orwellian plan to store Census Name and Address for years, 2016. Disponível em: https://privacy.org.au/Campaigns/Census2016/ABS_Census_MR_16.1.pdf.

[13] Autoriteit Persoonsgegevens. Autoriteit Persoonsgegevens. Disponível em: https://autoriteitpersoonsgegevens.nl.

[14] Autoriteit Persoonsgegevens. Gebruik van persoonsgegevens in het onderwijs. Disponível em: https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/onderwijs/gebruik-van-persoonsgegevens-het-onderwijs.

[15] Autoriteit Persoonsgegevens. Onderzoek naar de toegangsrechten van medewerkers van onderwijsinstellingen van ASKO tot persoonsgegevens van leerlingen in leerlingvolgsysteem X. 2018. Disponível em: https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/01_onderzoeksrapport_asko.pdf.

[16] BBC. Edward Snowden: Leaks that exposed US spy programme. *BBC*, 2014. Disponível em: https://www.bbc.com/news/world-us-canada-23123964.

[17] BCC. Australian census attacked by hackers. *BBC*, 2016. Disponível em: https://www.bbc.com/news/world-australia-37008173.

[18] Bruno Ricardo Bioni. Xeque-mate: o tripé da proteção de dados pessoais no jogo de xadrez das iniciativas legislativas no brasil. *GPoPAI/USP*, 2016.

[19] Rodolfo Borges. Do CPF na farmácia às redes: como nova lei protegerá seus dados pessoais. *El País*, 2018. Disponível em: https://brasil.elpais.com/brasil/2018/07/11/politica/1531325313_478217.html.

[20] Luíza Couto Chaves Brandão. O Marco Civil da Internet e a Proteção de Dados: diálogos com a LGPD. In *Proteção de dados pessoais: privacidade versus avanço tecnológico. Cadernos Adenauer XX, nº3*. Fundação Konrad Adanauer, 2019. Disponível em: https://www.kas.de/documents/265553/265602/Caderno+Adenauer+3+Schutz+von+pers%C3%B6nlichen+Daten.pdf/476709fc-b7dc-8430-12f1-ba21564cde06?version=1.0&t=1571685012573.

[21] United States Census Bureau. Disclosure Avoidance for Census 2010 and American Community Survey Five-year Tabular Data Products, 2009. Disponível em: https://www.census.gov/library/working-papers/2009/adrm/rrs2009-10.html.

[22] United States Census Bureau. Statistical Quality Standards, 2013. https://www.census.gov/content/dam/Census/about/about-the-bureau/policies_and_notices/quality/statistical-quality-standards/Quality_Standards.pdf.

[23] United States Census Bureau. 2010 Demonstration Data Products, 2019. Disponível em: https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2010-demonstration-data-products.html.

[24] United States Census Bureau. Disclosure Avoidance Techniques Used for the 1960 Through 2010 Decennial Censuses of Population and Housing Public Use Microdata Samples, 2019. Disponível em: https://www.census.gov/library/working-papers/2019/adrm/six-decennial-censuses-da.html.

[25] United States Census Bureau. Legacy Techniques and Current Research in Disclosure Avoidance at the U.S. Census Bureau, 2019. Disponível em https://www.census.gov/library/working-papers/2019/adrm/legacy-da-techniques.html.

[26] United States Census Bureau. Status Update on the 2020 Census Data Products Plan, 2019. Disponível em: https://www2.census.gov/cac/nac/meetings/2019-11/devine-hollingsworth-status-update-2020-data-products-plan.pdf.

[27] CEPAL. Código Regional de Buenas Prácticas en Estadísticas para América Latina y el Caribe, 2011. Disponível em: http://www.ine.gub.uy/buenas-practicas.

[28] Conselho da União Europeia. Diretiva 95/46/CE do Parlamento Europeu e do Conselho, de 24 de outubro de 1995, 1995. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:31995L0046&from=PT.

[29] Conselho da União Europeia. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016, 2016. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32016R0679.

[30] Organização das Nações Unidas. Fundamental Principles of Official Statistics (A/RES/68/261 from 29 January 2014), 2014. Disponível em: https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx.

[31] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Portaria 370 - Plano de Dados Abertos, 2016. Disponível em: http://download.inep.gov.br/institucional/legislacao/2016/portaria_n370.pdf.

[32] Instituto Brasileiro de Geografia e Estatística. Princípios Fundamentais das Estatísticas Oficiais. Disponível em https://www.ibge.gov.br/institucional/documentos-ibge.html?option=com_content&view=article&id=16148.

[33] Instituto Brasileiro de Geografia e Estatística. Código de Boas Práticas das Estatísticas do IBGE, 2013. Disponível em ftp://ftp.ibge.gov.br/Informacoes_Gerais_e_Referencia/Codigo_de_Boas_Praticas_das_Estatisticas_do_IBGE.pdf.

[34] Instituto Brasileiro de Geografia e Estatística. Confidencialidade no IBGE - procedimentos adotados na preservação do sigilo das informações individuais nas divulgações de resultados das operações estatísticas, 2018. Disponível em: http://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101636.

[35] Dienst Uitvoering Onderwijs. Open data van DUO. Disponível em: https://duo.nl/open_onderwijsdata/.

[36] Irit Dinur and Kobbi Nissim. Revealing information while preserving privacy. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 202–210, 2003.

[37] Governo do Brasil. Constituição Federal. Texto compilado até a Emenda Constitucional nº 105 de 12/12/2019. Disponível em: https://www.senado.leg.br/atividade/const/con1988/con1988_12.12.2019/ind.asp.

[38] Governo do Brasil. Decreto 6.317, de 20 de dezembro de 2007. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2007/Decreto/D6317.htm.

[39] Governo do Brasil. Decreto 6.425, de abril de 2008. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Decreto/D6425.htm.

[40] Governo do Brasil. Decreto 73.177, de 20 de novembro de 1973. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto/Antigos/D73177.htm.

[41] Governo do Brasil. Lei 13.709, de 14 de agosto de 2018, Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/Lei/L13709.htm.

[42] Governo do Brasil. Lei 14.010, de 10 de junho de 2020. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2020/Lei/L14010.htm.

[43] Governo do Brasil. Lei 5.534, de 14 de novembro de 1968. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/L5534.htm.

[44] Governo do Brasil. Lei 9.394, de dezembro de 1996. Disponível em: http://www.planalto.gov.br/ccivil_03/LEIS/L9394.htm.

[45] Governo do Brasil. Lei 9.448, de 14 de março de 1997. Disponível em: http://www.planalto.gov.br/ccivil_03/LEIS/L9448.htm.

[46] Governo do Brasil. Lei Complementar 131, de 27 de maio de 2009, Lei Capiberibe. Disponível em: http://www.planalto.gov.br/ccivil_03/LEIS/LCP/Lcp131.htm.

[47] Governo do Brasil. Lei de Acesso à Informação (LAI). Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12527.htm.

[48] Governo do Brasil. Portal da Transparência do Governo Federal, Controladoria-Geral da União. Disponível em: https://www.gov.br/cgu/pt-br/assuntos/transparencia-publica/portal-da-transparencia.

[49] Vikram Dodd. Met police to begin using live facial recognition cameras in London. *The Guardian*, 2020. Disponível em: https://www.theguardian.com/technology/2020/jan/24/met-police-begin-using-live-facial-recognition-cameras.

[50] Cynthia Dwork. A firm foundation for private data analysis. *Communications of the ACM*, 54(1):86–95, 2011.

[51] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

[52] ECK. Documentatie. Disponível em: https://www.eck-id.nl/implementatie/documentatie.

[53] EUROSTAT. Código de Conduta para as Estatísticas Europeias, 2017. Disponível em: https://ec.europa.eu/eurostat/documents/4031688/9332182/KS-02-18-142-PT-N.pdf/acea71f5-e1b1-4bcc-b4db-7cb98ea600dd.

[54] EUROSTAT. Quality Assurance Framework, 2019. Disponível em: https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646.

[55] Emily Feng. How China Is Using Facial Recognition Technology. *NPR*, 2019. Disponível em: https://www.npr.org/2019/12/16/788597818/how-china-is-using-facial-recognition-technology.

[56] Simson Garfinkel. *Database Nation: The Death of Privacy in the 21st Century*. O'Reilly & Associates, Inc., USA, 2000.

[57] Simson Garfinkel, John M Abowd, and Christian Martindale. Understanding database reconstruction attacks on public data. *Queue*, 16(5):28–53, 2018.

[58] Simson L Garfinkel, John M Abowd, and Sarah Powazek. Issues encountered deploying differential privacy. In *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, pages 133–137, 2018.

[59] Governo do Brasil. Lei 12.965, de 23 de abril de 2014, Marco Civil da Internet. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm.

[60] Governo do Brasil. Lei 9.507, de 12 de novembro de 1997. Disponível em http://www.planalto.gov.br/ccivil_03/LEIS/L9507.htm.

[61] Governo do Brasil. Medida Provisória 869, de 2018. Proteção de dados pessoais. Disponível em https://www.congressonacional.leg.br/materias/medidas-provisorias/-/mpv/135062.

[62] Governo do Brasil. Medida Provisória 954, de 17 de abril de 2020. Disponível em: http://www.planalto.gov.br/CCIVIL_03/_Ato2019-2022/2020/Mpv/mpv954.htm.

[63] Governo do Brasil. Medida Provisória 959, de 2020. Regras para o auxílio emergencial e adiamento da vigência da LGPD. Disponível em https://www.congressonacional.leg.br/materias/medidas-provisorias/-/mpv/141753.

[64] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Portaria 91, de 2 de fevereiro de 2017, 2017. Disponível em: http://download.inep.gov.br/educacao_basica/censo_escolar/legislacao/2017/portaria_inep_91_02022017_principios_fundamentais_estatisticas_eduacionais.pdf.

[65] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Portaria 492, de 7 de junho de 2018, 2018.

[66] Gabriel Joppert. O que diz a lei atual quando o governo deixa vazar nossos dados pessoais? *UOL*, 2019. Disponível em: https://www.uol.com.br/tilt/noticias/redacao/2019/10/16/vazamentos-comprometem-dados-de-dezenas-de-milhoes-de-brasileiros.htm.

[67] Alexis Kateifides, Joel Bates, Nikos Papageorgiou, Rumer Ramsey, Bart van der Geest, Alice Marini, and Pascale Arguinarena. Comparing privacy laws: GDPR v. LGPD, 2019. Disponível em: https://baptistaluz.com.br/wp-content/uploads/2019/05/DataGuidance-GPDR-LGPD-For-Print.pdf.

[68] Kennisnet. Nummervoorziening - Java Client Reference Application, 2018. Disponível em: https://github.com/kennisnet/Nummervoorziening-JavaReferenceImplementation.

[69] Kennisnet, Marc Fleischeuers, and Vincent Tedjakusuma. ECK iD Principes en processen, 2019. Disponível em: https://developers.wiki.kennisnet.nl/images/d/de/Principes_en_processen_ECK_ID.pdf.

[70] Richie Koch. LGPD: a versão brasileira do regulamento europeu. *Serviço Federal de Processamento de Dados*, 2019. Disponível em: https://www.serpro.gov.br/lgpd/noticias/lgpd-versao-brasileira-gdpr-dados-pessoais.

[71] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. L-Diversity: Privacy beyond k-Anonymity. *ACM Transactions on Knowledge Discovery from Data*, 1(1):3–es, March 2007.

[72] Angela Mari. Data privacy awareness grows in brazil. *ZDNet*, 2020. Disponível em: https://www.zdnet.com/article/data-privacy-awareness-grows-in-brazil/.

[73] Jeffrey Mervis. Can a set of equations keep U.S. census data private? *Science Magazine*, 2019. Disponível em: https://www.sciencemag.org/news/2019/01/can-set-equations-keep-us-census-data-private.

[74] Eduardo Militão. Falha de cartórios expõe dados de ao menos 1 milhão de pais, mães e filhos. *UOL*, 2019. Disponível em: https://noticias.uol.com.br/cotidiano/ultimas-noticias/2019/10/29/falha-de-cartorios-expoe-dados-de-ao-menos-1-milhao-de-pais-maes-e-filhos.htm.

[75] Ministério das Relações Exteriores do Brasil. O Brasil e a OCDE. Disponível em: http://www.itamaraty.gov.br/pt-BR/politica-externa/diplomacia-economica-comercial-e-financeira/15584-o-brasil-e-a-ocde.

[76] Arvind Narayanan and Vitaly Shmatikov. Robust De-anonymization of Large Sparse Datasets. In *Proc. of S&P*, pages 111–125, 2008.

[77] Netflix. Netflix Price. Disponível em: https://www.netflixprize.com/.

[78] Helen Nissenbaum. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, 2009.

[79] Government of Australia. Australian Bureau of Statistics Act 1975. Disponível em: https://www.legislation.gov.au/Details/C2017C00096.

[80] Government of Australia. Census and Statistics Act 1905. Disponível em: https://www.legislation.gov.au/Details/C2016C01005.

[81] Government of Australia. Privacy Act 1988. Disponível em: https://www.legislation.gov.au/Details/C2015C00598.

[82] Australian Bureau of Statistics. 2016 Census Privacy Policy. Disponível em: https://www.abs.gov.au/websitedbs/censushome.nsf/home/privacypolicy?opendocument&navpos=110.

[83] Australian Bureau of Statistics. Australian Census Longitudinal Dataset. Disponível em: https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2011.0.55.001~2016~Main%20Features~Australian%20Census%20Longitudinal%20Dataset~12.

[84] Australian Bureau of Statistics. Census Microdata. Disponível em: https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2011.0.55.001~2016~Main%20Features~Census%20Microdata~13.

[85] Australian Bureau of Statistics. Microdata prices. Disponível em: https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Microdata+prices.

[86] Australian Bureau of Statistics. Privacy, Confidentiality & Security. Disponível em: https://www.abs.gov.au/websitedbs/censushome.nsf/home/privacy?opendocument&navpos=130.

[87] Government of the United States of America. Statistical Standards Program - Confidentiality Laws. Disponível em: https://nces.ed.gov/statprog/conflaws.asp.

[88] Government of the United States of America. U.S. Code, Title 13. Disponível em: https://www.law.cornell.edu/uscode/text/13.

[89] Government of the United States of America. U.S. Code, Title 15, Section 552a. Disponível em: https://www.law.cornell.edu/uscode/text/5/552a.

[90] Government of the United States of America. U.S. Code, Title 15, section 552a. Disponível em: https://www.law.cornell.edu/uscode/text/15.

[91] Government of the United States of America. USA PATRIOT Act of 2001, 2001. Disponível em https://www.fincen.gov/resources/statutes-regulations/usa-patriot-act.

[92] Government of the United States of America. Confidential Information Protection and Statistical Efficiency Act (CIPSEA), 2002. Disponível em: https://www.eia.gov/cipsea/cipsea.pdf.

[93] Government of the United States of America. Statistical Policy Working Paper 22 - Report on Statistical Disclosure Limitation Methodology, 2005. Disponível em: https://nces.ed.gov/FCSM/pdf/spwp22.pdf.

[94] Government of the United States of America. 2012 Revision of the National Center for Education Statistics Statistical Standards, 2012. Disponível em: https://nces.ed.gov/statprog/2012/.

[95] Government of the United States of America. Overview of the Privacy Act of 1974, 2015. Disponível em: https://www.justice.gov/opcl/overview-privacy-act-1974-2015-edition.

[96] Organização para a Cooperação e Desenvolvimento Econômico - OCDE. OECD privacy guidelines. Disponível em: https://www.oecd.org/internet/ieconomy/privacy-guidelines.htm.

[97] Organização para a Cooperação e Desenvolvimento Econômico - OCDE. Growing Up Online: Addressing the Needs of Children in the Digital Environment, 2020. Disponível em: https://www.oecd.org/going-digital/growing-up-online.pdf.

[98] Daniel Palmer. ABS to permanently store personal data from Australian census. *Delimiter*, 2016. Disponível em: https://delimiter.com.au/2016/02/01/abs-permanently-store-personal-data-australian-census/.

[99] Filipe Pontes. RGPD e os seus direitos. *Visão*, 2018. Disponível em: https://visao.sapo.pt/opiniao/ponto-de-vista/silencio-da-fraude/2018-05-30-rgpd-e-os-seus-direitos/.

[100] Fabian Prasser, Florian Kohlmayer, Ronald Lautenschlaeger, and Klaus A Kuhn. Arx-a comprehensive tool for anonymizing biomedical data. In *AMIA Annual Symposium Proceedings*, volume 2014, page 984. American Medical Informatics Association, 2014.

[101] Privacyconvenant Onderwijs. Privacyconvenant Onderwijs. Disponível em: https://www.privacyconvenant.nl/.

[102] MJ Queiroz and GHMB Motta. Privacidade e Transparência no Setor público: Um Estudo de Caso da Publicação de Microdados do INEP. In *XV Simposio Brasileiro em Seguranca da Informacao e de Sistemas Computacionais-SBSeg*, 2015.

[103] Clarie Reilly. The census wasn't hacked, but australia still has a problem. *CNET*, 2016. Disponível em: https://www.cnet.com/news/the-census-wasnt-hacked-but-the-abs-still-has-a-problem/.

[104] Elettra Ronchi and Lisa Robinson. Child protection online. In *Educating 21st Century Children: Emotional Well-being in the Digital Age*. Organização para a Cooperação e Desenvolvimento Econômico - OCDE, 2019. Disponível em: https://www.oecd-ilibrary.org/sites/796ac574-en/index.html?itemId=/content/component/796ac574-en&mimeType=text/.

[105] Pierangela Samarati and Latanya Sweeney. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression, 1998.

[106] Search-ID. AOL user #4417749: Thelma Arnold. Disponível em: http://explicit-id.com/user/4417749-thelma_arnold.

[107] Paula Soprana. Governo vai usar dados de operadoras para monitorar aglomeração na pandemia. *Folha de S.Paulo*, 2020. Disponível em: https://www1.folha.uol.com.br/mercado/2020/04/governo-vai-usar-dados-de-operadoras-para-monitorar-deslocamentos-na-pandemia.shtml.

[108] Alice Stollmeyer, Marietje Schaake, and Frank Dignum. The Dutch tracing app 'soap opera' - lessons for Europe. *EUobserver*, 2020. Disponível em: https://euobserver.com/opinion/148265.

[109] Latanya Sweeney. Simple Demographics Often Identify People Uniquely, 2000. Disponível em: https://kilthub.cmu.edu/articles/Simple_Demographics_Often_Identify_People_Uniquely/6625769/1.

Anexos

A Constituição da República Federativa do Brasil: Principais trechos relacionados ao escopo deste projeto

A seguir apresentamos uma seleção dos trechos da Constituição da República Federativa do Brasil [37] identificados como os mais relevantes no escopo do presente TED 8750.

> **Art. 5º** Todos são iguais perante a lei, sem distinção de qualquer natureza, garantindo-se aos brasileiros e aos estrangeiros residentes no País a inviolabilidade do direito à vida, à liberdade, à igualdade, à segurança e à propriedade, nos termos seguintes:
>
> X são invioláveis a intimidade, a vida privada, a honra e a imagem das pessoas, assegurado o direito a indenização pelo dano material ou moral decorrente de sua violação;
>
> XIV é assegurado a todos o acesso à informação e resguardado o sigilo da fonte, quando necessário ao exercício profissional;
>
> XXXIII todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado;
>
> LXXII conceder-se-á *habeas data*:
>
> a) para assegurar o conhecimento de informações relativas à pessoa do impetrante, constantes de registros ou bancos de dados de entidades governamentais ou de caráter público;
>
> b) para a retificação de dados, quando não se prefira fazê-lo por processo sigiloso, judicial ou administrativo;
>
> LXXVII são gratuitas as ações de habeas corpus e habeas data, e, na forma da lei, os atos necessários ao exercício da cidadania.

Art. 60. A Constituição poderá ser emendada mediante proposta:

§ 4º Não será objeto de deliberação a proposta de emenda tendente a abolir:

IV os direitos e garantias individuais.

B LAI: Principais trechos relacionados ao escopo deste projeto

A seguir apresentamos uma seleção dos trechos da Lei 12.527/2011, conhecida como *Lei de Acesso à Informação* (LAI) [47], identificados como os mais relevantes no escopo do presente TED 8750.

> **Art. 1º** Esta Lei dispõe sobre os procedimentos a serem observados pela União, Estados, Distrito Federal e Municípios, com o fim de garantir o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 § 2º do art. 216 da Constituição Federal [37].
>
> Parágrafo único. Subordinam-se ao regime desta Lei:
>
> I os órgãos públicos integrantes da administração direta dos Poderes Executivo, Legislativo, incluindo as Cortes de Contas, e Judiciário e do Ministério Público;
>
> II as autarquias, as fundações públicas, as empresas públicas, as sociedades de economia mista e demais entidades controladas direta ou indiretamente pela União, Estados, Distrito Federal e Municípios.
>
> **Art. 3º** Os procedimentos previstos nesta Lei destinam-se a assegurar o direito fundamental de acesso à informação e devem ser executados em conformidade com os princípios básicos da administração pública e com as seguintes diretrizes:
>
> I observância da publicidade como preceito geral e do sigilo como exceção;
>
> II divulgação de informações de interesse público, independentemente de solicitações;
>
> III utilização de meios de comunicação viabilizados pela tecnologia da informação;
>
> IV fomento ao desenvolvimento da cultura de transparência na administração pública;

V desenvolvimento do controle social da administração pública.

Art. 4º Para os efeitos desta Lei, considera-se:

I informação: dados, processados ou não, que podem ser utilizados para produção e transmissão de conhecimento, contidos em qualquer meio, suporte ou formato;

II documento: unidade de registro de informações, qualquer que seja o suporte ou formato;

III informação sigilosa: aquela submetida temporariamente à restrição de acesso público em razão de sua imprescindibilidade para a segurança da sociedade e do Estado;

IV informação pessoal: aquela relacionada à pessoa natural identificada ou identificável;

V tratamento da informação: conjunto de ações referentes à produção, recepção, classificação, utilização, acesso, reprodução, transporte, transmissão, distribuição, arquivamento, armazenamento, eliminação, avaliação, destinação ou controle da informação;

VI disponibilidade: qualidade da informação que pode ser conhecida e utilizada por indivíduos, equipamentos ou sistemas autorizados;

VII autenticidade: qualidade da informação que tenha sido produzida, expedida, recebida ou modificada por determinado indivíduo, equipamento ou sistema;

VIII integridade: qualidade da informação não modificada, inclusive quanto à origem, trânsito e destino;

IX primariedade: qualidade da informação coletada na fonte, com o máximo de detalhamento possível, sem modificações.

Art. 6º Cabe aos órgãos e entidades do poder público, observadas as normas e procedimentos específicos aplicáveis, assegurar a:

I gestão transparente da informação, propiciando amplo acesso a ela e sua divulgação;

II proteção da informação, garantindo-se sua disponibilidade, autenticidade e integridade; e

III proteção da informação sigilosa e da informação pessoal, observada a sua disponibilidade, autenticidade, integridade e eventual restrição de acesso.

Art. 7º O acesso à informação de que trata esta Lei compreende, entre outros, os direitos de obter:

I orientação sobre os procedimentos para a consecução de acesso, bem como sobre o local onde poderá ser encontrada ou obtida a informação almejada;

II informação contida em registros ou documentos, produzidos ou acumulados por seus órgãos ou entidades, recolhidos ou não a arquivos públicos;

III informação produzida ou custodiada por pessoa física ou entidade privada decorrente de qualquer vínculo com seus órgãos ou entidades, mesmo que esse vínculo já tenha cessado;

IV informação primária, íntegra, autêntica e atualizada;

§ 1º O acesso à informação previsto no caput não compreende as informações referentes a projetos de pesquisa e desenvolvimento científicos ou tecnológicos cujo sigilo seja imprescindível à segurança da sociedade e do Estado.

§ 2º Quando não for autorizado acesso integral à informação por ser ela parcialmente sigilosa, é assegurado o acesso à parte não sigilosa por meio de certidão, extrato ou cópia com ocultação da parte sob sigilo.

Art. 8º É dever dos órgãos e entidades públicas promover, independentemente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas ou custodiadas.

§ 2º Para cumprimento do disposto no **caput**, os órgãos e entidades públicas deverão utilizar todos os meios e instrumentos legítimos de que dispuserem, sendo obrigatória a divulgação em sítios oficiais da rede mundial de computadores (Internet).

§ 3º Os sítios de que trata o § 2º deverão, na forma de regulamento, atender, entre outros, aos seguintes requisitos:

I conter ferramenta de pesquisa de conteúdo que permita o acesso à informação de forma objetiva, transparente, clara e em linguagem de fácil compreensão;

II possibilitar a gravação de relatórios em diversos formatos eletrônicos, inclusive abertos e não proprietários, tais como planilhas e texto, de modo a facilitar a análise das informações;

III possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina;

IV divulgar em detalhes os formatos utilizados para estruturação da informação;

V garantir a autenticidade e a integridade das informações disponíveis para acesso;

VI manter atualizadas as informações disponíveis para acesso;

Art. 22. O disposto nesta Lei não exclui as demais hipóteses legais de sigilo e de segredo de justiça nem as hipóteses de segredo industrial decorrentes da exploração direta de atividade econômica pelo Estado ou por pessoa física ou entidade privada que tenha qualquer vínculo com o poder público.

Art. 31. O tratamento das informações pessoais deve ser feito de forma transparente e com respeito à intimidade, vida privada, honra e imagem das pessoas, bem como às liberdades e garantias individuais.

§ 1º As informações pessoais, a que se refere este artigo, relativas à intimidade, vida privada, honra e imagem:

I terão seu acesso restrito, independentemente de classificação de sigilo e pelo prazo máximo de 100 (cem) anos a contar da sua data de produção, a agentes públicos legalmente autorizados e à pessoa a que elas se referirem; e

II poderão ter autorizada sua divulgação ou acesso por terceiros diante de previsão legal ou consentimento expresso da pessoa a que elas se referirem.

§ 2º Aquele que obtiver acesso às informações de que trata este artigo será responsabilizado por seu uso indevido.

§ 3º O consentimento referido no inciso II do § 1º não será exigido quando as informações forem necessárias:

II à realização de estatísticas e pesquisas científicas de evidente interesse público ou geral, previstos em lei, sendo vedada a identificação da pessoa a que as informações se referirem;

V à proteção do interesse público e geral preponderante.

§ 5º Regulamento disporá sobre os procedimentos para tratamento de informação pessoal.

Art. 32. Constituem condutas ilícitas que ensejam responsabilidade do agente público ou militar:

I recusar-se a fornecer informação requerida nos termos desta Lei, retardar deliberadamente o seu fornecimento ou fornecê-la intencionalmente de forma incorreta, incompleta ou imprecisa;

II utilizar indevidamente, bem como subtrair, destruir, inutilizar, desfigurar, alterar ou ocultar, total ou parcialmente, informação que se encontre sob sua guarda ou a que tenha acesso ou conhecimento em razão do exercício das atribuições de cargo, emprego ou função pública;

IV divulgar ou permitir a divulgação ou acessar ou permitir acesso indevido à informação sigilosa ou informação pessoal;

Art. 38. Aplica-se, no que couber, a Lei nº 9.507, de 12 de novembro de 1997 [60], em relação à informação de pessoa, física ou jurídica, constante de registro ou banco de dados de entidades governamentais ou de caráter público.

C LGPD: Principais trechos relacionados ao escopo deste projeto

A seguir apresentamos uma seleção dos trechos da Lei Lei 13.709/2018, conhecida como *Lei Geral de Proteção de Dados Pessoais* (LGPDP ou LGPD) [41], identificados como os mais relevantes no escopo do presente TED 8750.

> **Art. 1º** Esta Lei dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural.
>
> Parágrafo único. As normas gerais contidas nesta Lei são de interesse nacional e devem ser observadas pela União, Estados, Distrito Federal e Municípios.
>
> **Art. 2º** A disciplina da proteção de dados pessoais tem como fundamentos:
>
> I o respeito à privacidade;
>
> II a autodeterminação informativa;
>
> III a liberdade de expressão, de informação, de comunicação e de opinião;
>
> IV a inviolabilidade da intimidade, da honra e da imagem;
>
> **Art. 3º** Esta Lei aplica-se a qualquer operação de tratamento realizada por pessoa natural ou por pessoa jurídica de direito público ou privado, independentemente do meio, do país de sua sede ou do país onde estejam localizados os dados, desde que:
>
> I a operação de tratamento seja realizada no território nacional;
>
> II a atividade de tratamento tenha por objetivo a oferta ou o fornecimento de bens ou serviços ou o tratamento de dados de indivíduos localizados no território nacional; ou
>
> III os dados pessoais objeto do tratamento tenham sido coletados no território nacional.

§ 1º Consideram-se coletados no território nacional os dados pessoais cujo titular nele se encontre no momento da coleta.

Art. 5º Para os fins desta Lei, considera-se:

I dado pessoal: informação relacionada a pessoa natural identificada ou identificável;

II dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural;

III dado anonimizado: dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento;

IV banco de dados: conjunto estruturado de dados pessoais, estabelecido em um ou em vários locais, em suporte eletrônico ou físico;

V titular: pessoa natural a quem se referem os dados pessoais que são objeto de tratamento;

VI controlador: pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados pessoais;

VII operador: pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador;

IX agentes de tratamento: o controlador e o operador;

X tratamento: toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração;

XI anonimização: utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo;

XVII relatório de impacto à proteção de dados pessoais: documentação do controlador que contém a descrição dos processos de tratamento de dados pessoais que podem gerar riscos às liberdades civis e aos direitos fundamentais, bem como medidas, salvaguardas e mecanismos de mitigação de risco;

XVIII órgão de pesquisa: órgão ou entidade da administração pública direta ou indireta ou pessoa jurídica de direito privado sem fins lucrativos legalmente constituída sob as leis brasileiras, com sede e foro no País, que inclua em sua missão institucional ou em seu objetivo social ou estatutário a pesquisa básica ou aplicada de caráter histórico, científico, tecnológico ou estatístico; e

XIX autoridade nacional: órgão da administração pública responsável por zelar, implementar e fiscalizar o cumprimento desta Lei em todo o território nacional.

Art. 6º As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios:

I finalidade: realização do tratamento para propósitos legítimos, específicos, explícitos e informados ao titular, sem possibilidade de tratamento posterior de forma incompatível com essas finalidades;

II adequação: compatibilidade do tratamento com as finalidades informadas ao titular, de acordo com o contexto do tratamento;

III necessidade: limitação do tratamento ao mínimo necessário para a realização de suas finalidades, com abrangência dos dados pertinentes, proporcionais e não excessivos em relação às finalidades do tratamento de dados;

IV livre acesso: garantia, aos titulares, de consulta facilitada e gratuita sobre a forma e a duração do tratamento, bem como sobre a integralidade de seus dados pessoais;

V qualidade dos dados: garantia, aos titulares, de exatidão, clareza, relevância e atualização dos dados, de acordo com a necessidade e para o cumprimento da finalidade de seu tratamento;

VI transparência: garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial;

VII segurança: utilização de medidas técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou difusão;

VIII prevenção: adoção de medidas para prevenir a ocorrência de danos em virtude do tratamento de dados pessoais;

IX não discriminação: impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos;

X responsabilização e prestação de contas: demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de dados pessoais e, inclusive, da eficácia dessas medidas.

Art. 7º O tratamento de dados pessoais somente poderá ser realizado nas seguintes hipóteses:

I mediante o fornecimento de consentimento pelo titular;

II para o cumprimento de obrigação legal ou regulatória pelo controlador;

III pela administração pública, para o tratamento e uso compartilhado de dados necessários à execução de políticas públicas previstas em leis e regulamentos ou respaldadas em contratos, convênios ou instrumentos congêneres, observadas as disposições do Capítulo IV desta Lei;

IV para a realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais;

IX quando necessário para atender aos interesses legítimos do controlador ou de terceiro, exceto no caso de prevalecerem direitos e liberdades fundamentais do titular que exijam a proteção dos dados pessoais;

§ 3º O tratamento de dados pessoais cujo acesso é público deve considerar a finalidade, a boa-fé e o interesse público que justificaram sua disponibilização.

Art. 11. O tratamento de dados pessoais sensíveis somente poderá ocorrer nas seguintes hipóteses:

I quando o titular ou seu responsável legal consentir, de forma específica e destacada, para finalidades específicas;

II sem fornecimento de consentimento do titular, nas hipóteses em que for indispensável para:

a) cumprimento de obrigação legal ou regulatória pelo controlador;

b) tratamento compartilhado de dados necessários à execução, pela administração pública, de políticas públicas previstas em leis ou regulamentos;

c) realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais sensíveis;

§ 1º Aplica-se o disposto neste artigo a qualquer tratamento de dados pessoais que revele dados pessoais sensíveis e que possa causar dano ao titular, ressalvado o disposto em legislação específica.

§ 2º Nos casos de aplicação do disposto nas alíneas “a” e “b” do inciso II do caput deste artigo pelos órgãos e pelas entidades públicas, será dada publicidade à referida dispensa de consentimento, nos termos do inciso I do caput do art. 23 desta Lei.

Art. 12. Os dados anonimizados não serão considerados dados pessoais para os fins desta Lei, salvo quando o processo de anonimização ao qual foram submetidos for revertido, utilizando exclusivamente meios próprios, ou quando, com esforços razoáveis, puder ser revertido.

§ 1º A determinação do que seja razoável deve levar em consideração fatores objetivos, tais como custo e tempo necessários para reverter o processo de anonimização, de acordo com as tecnologias disponíveis, e a utilização exclusiva de meios próprios.

§ 2º Poderão ser igualmente considerados como dados pessoais, para os fins desta Lei, aqueles utilizados para formação do perfil comportamental de determinada pessoa natural, se identificada.

§ 3º A autoridade nacional poderá dispor sobre padrões e técnicas utilizados em processos de anonimização e realizar verificações acerca de sua segurança, ouvido o Conselho Nacional de Proteção de Dados Pessoais.

Art. 13. Na realização de estudos em saúde pública, os órgãos de pesquisa poderão ter acesso a bases de dados pessoais, que serão tratados exclusivamente dentro do órgão e estritamente para a finalidade de realização de estudos e pesquisas e mantidos em ambiente controlado e seguro, conforme práticas de segurança previstas em regulamento específico e que incluam, sempre que possível, a anonimização ou pseudonimização dos dados, bem como considerem os devidos padrões éticos relacionados a estudos e pesquisas.

§ 4º Para os efeitos deste artigo, a pseudonimização é o tratamento por meio do qual um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo, senão pelo uso de informação adicional mantida separadamente pelo controlador em ambiente controlado e seguro.

Art. 14. O tratamento de dados pessoais de crianças e de adolescentes deverá ser realizado em seu melhor interesse, nos termos deste artigo e da legislação pertinente.

§ 1º O tratamento de dados pessoais de crianças deverá ser realizado com o consentimento específico e em destaque dado por pelo menos um dos pais ou pelo responsável legal.

§ 2º No tratamento de dados de que trata o § 1º deste artigo, os controladores deverão manter pública a informação sobre os tipos de dados coletados, a forma de sua utilização e os procedimentos para o exercício dos direitos a que se refere o art. 18 desta Lei.

§ 3º Poderão ser coletados dados pessoais de crianças sem o consentimento a que se refere o § 1º deste artigo quando a coleta for necessária para contatar os pais ou o responsável legal, utilizados uma única vez e sem armazenamento, ou para sua proteção, e em nenhum caso poderão ser repassados a terceiro sem o consentimento de que trata o § 1º deste artigo.

§ 4º Os controladores não deverão condicionar a participação dos titulares de que trata o § 1º deste artigo em jogos, aplicações de Internet ou outras atividades ao fornecimento de informações pessoais além das estritamente necessárias à atividade.

§ 5º O controlador deve realizar todos os esforços razoáveis para verificar que o consentimento a que se refere o § 1º deste artigo foi dado pelo responsável pela criança, consideradas as tecnologias disponíveis.

§ 6º As informações sobre o tratamento de dados referidas neste artigo deverão ser fornecidas de maneira simples, clara e acessível, consideradas as

características físico-motoras, perceptivas, sensoriais, intelectuais e mentais do usuário, com uso de recursos audiovisuais quando adequado, de forma a proporcionar a informação necessária aos pais ou ao responsável legal e adequada ao entendimento da criança.

Art. 16. Os dados pessoais serão eliminados após o término de seu tratamento, no âmbito e nos limites técnicos das atividades, autorizada a conservação para as seguintes finalidades:

I cumprimento de obrigação legal ou regulatória pelo controlador;

II estudo por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais;

III transferência a terceiro, desde que respeitados os requisitos de tratamento de dados dispostos nesta Lei; ou

IV uso exclusivo do controlador, vedado seu acesso por terceiro, e desde que anonimizados os dados.

Art. 17. Toda pessoa natural tem assegurada a titularidade de seus dados pessoais e garantidos os direitos fundamentais de liberdade, de intimidade e de privacidade, nos termos desta Lei.

D Principais diferenças entre a LGPD brasileira e a GDPR europeia

A Lei Geral de Proteção de Dados Pessoais (LGPD) brasileira (discutida na Seção 3.1) foi fortemente inspirada pelo seu equivalente europeu, o Regulamento Geral de Proteção de Dados (RGPD ou GDPR) (discutido na Seção 2.3.1), com o qual compartilha filosofia e abrangência similares. De fato, pode-se constatar que a legislação brasileira relativa à privacidade se aproxima mais da europeia, que por sua vez é muito diferente da legislação dos EUA (discutida na Seção 2.2.1). Entretanto, há significativas diferenças entre a LGPD e o GDPR.

Uma comparação rica em detalhes, possivelmente do interesse de profissionais da área jurídica, pode ser encontrada no relatório criado em conjunto pela equipe do *DataGuidance by OneTrust* e do escritório Baptista Luz Advogados [67].

A seguir citamos o paralelo abaixo oferecido por Richie Koch, editor-chefe do portal GDPR.EU, [1] retirado de seu artigo sobre o assunto [70].

> **Diferenças entre a LGPD e o GDPR** Apesar de seus objetivos similares e a aparente influência do GDPR sobre os legisladores brasileiros, existem algumas diferenças marcantes a serem notadas entre as duas legislações.
>
> **Encarregados de proteção de dados** Ambos os textos legais demandam que os negócios e as organizações contratem um encarregado de proteção de dados. No entanto, enquanto a GDPR define quando um encarregado é necessário, o artigo 41 da LGPD diz, simplesmente: "O controlador deverá indicar encarregado pelo tratamento de dados pessoais", o que sugere que qualquer organização que processa dados de pessoas no Brasil terá de contratar alguém para tal posto. Esta é outra área que provavelmente receberá maior clarificação, mas da forma como está escrito, é uma das áreas onde a LGPD é mais rígida que a GDPR.

[1] https://gdpr.eu/

Base legal para processamento de dados Possivelmente a mais significativa diferença entre a LGPD e o GDPR recai sobre o que se qualifica como base legal para processar dados. O GDPR tem seis bases legais para o processamento, e um controlador de dados deve escolher uma delas como justificativa para utilizar a informação de um titular de dados. No entanto, no seu artigo 7º, a LGPD lista dez. São eles:

- Mediante o fornecimento de consentimento pelo titular;
- Para o cumprimento de obrigação legal ou regulatória pelo controlador;
- Pela administração pública, para o tratamento e uso compartilhado de dados necessários à execução de políticas públicas previstas em leis e regulamentos ou respaldadas em contratos, convênios ou instrumentos congêneres, observadas as disposições do capítulo IV da Lei;
- Para a realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais;
- Quando necessário para a execução de contrato ou de procedimentos preliminares relacionados a contrato do qual seja parte o titular, a pedido do titular dos dados;
- Para o exercício regular de direitos em processo judicial, administrativo ou arbitral, esse último nos termos da Lei nº 9.307, de 23 de setembro de 1996 (*Lei de Arbitragem*);
- Para a proteção da vida ou da incolumidade física do titular ou de terceiro;
- Para a tutela da saúde, exclusivamente, em procedimento realizado por profissionais de saúde, serviços de saúde ou autoridade sanitária;
- Quando necessário para atender aos interesses legítimos do controlador ou de terceiro, exceto no caso de prevalecerem direitos e liberdades fundamentais do titular que exijam a proteção dos dados pessoais;
- Para a proteção do crédito, inclusive quanto ao disposto na legislação pertinente.

Ter a proteção do crédito como base legal para o processamento de dados é de fato uma diferença substancial em relação ao GDPR.

Avisos sobre quebra de sigilo de dados Embora tanto o GDPR quanto a LGPD requeiram que as organizações avisem sobre quebra de sigilo de dados à autoridade nacional de proteção de dados, o nível de especificidade varia amplamente entre as duas leis. O GDPR é explícito: uma organização deve avisar sobre qualquer quebra de sigilo em até 72 horas a partir de sua descoberta (embora diferentes organizações já estejam testando este limite).

A LGPD não oferece nenhum limite definido: o artigo 48 simplesmente diz que "O controlador deverá comunicar à autoridade nacional e ao titular a ocorrência de incidente de segurança que possa acarretar risco ou dano relevante aos titulares [...] em prazo razoável, conforme definido pela autoridade nacional". Uma vez que a agência nacional de proteção de dados não foi, até o momento, estabelecida, não há indicação sobre o que constitui "prazo razoável".

Multas Uma lei é tão forte quanto seus "dentes". É por isso que as multas máximas do GDPR são substanciais, obrigando organizações que cometem violações graves da legislação europeia a pagarem até 20 milhões de euros, ou 4% de sua receita global anual, o que for maior.

As multas da LGPD são muito menos severas. O artigo 52 define que a multa máxima para uma violação é "de até 2% (dois por cento) do faturamento da pessoa jurídica de direito privado, grupo ou conglomerado no Brasil no seu último exercício, excluídos os tributos, limitada, no total, a R$ 50 milhões de reais por infração", ou cerca de 11 milhões de euros. As multas da LGPD estão em linha com as multas da GDPR para infrações menores, mas 11 milhões de euros não vão preocupar os maiores processadores de dados do mundo.

E Um panorama da percepção da questão de privacidade pela sociedade brasileira

Desde a aprovação da Lei Geral de Proteção de Dados Pessoais (LGPDP ou LGPD) pelo Senado Federal em 2018, a mídia brasileira vem cobrindo com atenção crescente temas relacionados à privacidade, especialmente quando se trata de dados coletados pelo Estado. Ainda em 2018, em comentário feito à edição brasileira do jornal El País [19], Flávia Lefèvre, conselheira da ONG Proteste, destacou o papel dos poderes públicos frente à proteção de dados pessoais. De acordo com Lefèvre, órgãos públicos são responsáveis pela coleta de muito mais dados do que empresas privadas, uma vez que processam, por exemplo, dados referentes a programas sociais, a declarações de imposto de renda, e a programas de saúde. Por esse motivo, a coleta e o tratamento de dados pelo Estado foram explicitamente incluídos na redação da LGPD.

Já em 2019, a questão da privacidade voltou a ser tratada pela imprensa com o vazamento de dados de aproximadamente 70 milhões de pessoas pelo Detran do Rio Grande do Norte, e com a descoberta de uma base de dados com informações sobre 92 milhões de brasileiros à venda na Internet. Dentre as análises jurídicas publicadas em reportagem do UOL [66], vale destacar aquela do advogado Diego Borba, consultor de segurança digital da Nielsen. De acordo com Borba, caso a LGPD já estivesse em vigor à época, a aplicação das penalizações previstas dependeria, por exemplo, dos métodos de prevenção adotados pelos órgãos, assim como das políticas e dos protocolos de segurança em caso de incidentes.

Outro incidente ocorrido em 2019 foi a exposição de informações de ao menos um milhão de brasileiros devido a uma falha de segurança que expôs milhares de arquivos de cartórios de São Paulo à Internet. Apesar de tratar-se de informações públicas, o acesso às mesmas deve ser controlado e regras e restrições aplicadas caso a caso, conforme declaração do tabelião e ex-delegado da Polícia Federal, Sandro Ferreira, ao UOL [74].

Já em 2020, com a chegada da pandemia de COVID-19 ao Brasil, foi anunciado pelo Ministério de Ciência e Tecnologia um acordo com grandes operadoras de telefonia móvel com o objetivo de monitorar aglomerações. Apesar de a LGPD ainda não estar em

vigor, as operadoras anunciaram que a plataforma criada para a colaboração respeita as normas da Lei, além daquelas estabelecidas pelo Marco Civil da Internet [59]. Para tanto, os dados fornecidos ao governo seriam organizados de forma agregada, estatística, e anônima. Entretanto, conforme apontado por Bruno Bioni, fundador do Data Privacy Brasil em comentário ao jornal Folha de São Paulo [107], é necessário que as medidas adotadas sejam divulgadas para que o cidadão e outros agentes possam proceder ao escrutínio público.

Outra iniciativa do governo federal frente à pandemia de COVID-19 foi estipulada pela Medida Provisória 954 [62], de abril de 2020, que permitiu o compartilhamento de informações cadastrais de usuários de linhas telefônicas com o IBGE. Entretanto, a Medida Provisória foi suspensa pelo Supremo Tribunal Federal por desrespeitar o direito à intimidade e ao sigilo da vida privada dos usuários [9], uma vez que os microdados seriam compartilhados sem qualquer tratamento prévio.

Finalmente, uma pesquisa realizada pela empresa de segurança Kaspersky e reportada pelo site de notícias ZDNet [72] aponta que 74% dos brasileiros entrevistados já tentaram remover informações pessoais de páginas na Internet ou redes sociais. Além disso, 58% tentam esconder informações pessoais ao navegarem pela Internet de modo a se protegerem de criminosos, enquanto 26% se dizem muito preocupados com as informações pessoais coletadas por aplicativos instalados em seus dispositivos.

Lista de Siglas

ABS *Australian Bureau of Statistics*.

ACLD *Australian Census Longitudinal Dataset*.

AHRQ *Agency for Healthcare Research & Quality*.

ANPD Autoridade Nacional de Proteção de Dados.

AP Autoriteit Persoonsgegevens (Autoridade de Privacidade holandesa).

ASKO *Amsterdamse Stichting voor Katholiek, Protestants-Christelijk en Interconfessioneel Onderwijs* (Fundação de Amsterdã para a Educação Católica, Cristã Protestante e Interconfessional).

AVG *Algemene Verordening Gegevensbescherming* (Lei holandesa que implementa a GDPR europeia).

BEA *Bureau of Economic Analysis*.

BJS *Bureau of Justice Statistics*.

BLS *Bureau of Labor Statistics*.

BTS *Bureau of Transportation Statistics*.

CEA Conferência Estatística das Américas.

CEPAL Comissão Econômica para a América Latina e o Caribe.

DCC Departamento de Ciência da Computação da UFMG.

DEED Diretoria de Estatísticas Educacionais do Inep.

DPIA *Data Protection Impact Assessment*.

DRB *Disclosure Review Board*.

ED *United States Department of Education*.

EEE Espaço Econômico Europeu.

EIA *Energy Information Administration*.

ENADE Exame Nacional de Desempenho dos Estudantes.

ENCCEJA Exame Nacional de Certificação de Competências de Jovens e Adultos.

ENEM Exame Nacional do Ensino Médio.

ERS *Economic Research Service*.

EUA Estados Unidos da América.

EUROSTAT Serviço de Estatística da União Europeia.

GDPR *General Data Protection Regulation*.

IBGE Instituto Brasileiro de Geografia e Estatística.

IES Instituição de Ensino Superior.

Inep Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.

IRIS Instituto de Referência em Internet e Sociedade.

IRS *Internal Revenue Service, Statistics of Income Division*.

LAI Lei de Acesso à Informação.

LGPD Lei Geral de Proteção de Dados Pessoais.

LGPDP Lei Geral de Proteção de Dados Pessoais.

MCI Marco Civil da Internet.

MEC Ministério da Educação.

NASS *National Agricultural Statistics Service*.

NCES *National Center for Education Statistics*.

NCHS *National Center for Health Statistics*.

NSF *National Science Foundation*.

OCDE Organização para a Cooperação e Desenvolvimento Econômico.

ONU Organização das Nações Unidas.

PUF *Public Use Files*.

PUMS *Public Use Microdata Samples*.

RDC *Research Data Centers*.

RGPD Regulamento Geral de Proteção de Dados.

SAEB Sistema de Avaliação da Educação Básica.

SAR Sala de Acesso a Dados Restritos.

SSA *Social Security Administration*.

TED Termo de Execução Desentralizada.

UE União Europeia.

UFMG Universidade Federal de Minas Gerais.

USCB *United States Census Bureau*.

Glossário

***Agency for Healthcare Research & Quality* (AHRQ)** Agência federal dos EUA responsável por divulgar dados sobre qualidade, adequação e eficácia dos serviços de saúde, assim como o acesso aos mesmos.

anonimização Utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo. Técnicas de anonimização incluem, mas não se limitam a, *desidentificação*, *k-anonymity*, *l-diversity* e *privacidade diferencial*, dentre outras. Diferentes técncias têm diferentes graus de sucesso, e algumas são comprovadamente menos eficazes que outras.

ASKO *Amsterdamse Stichting voor Katholiek, Protestants-Christelijk en Interconfessioneel Onderwijs* (Fundação de Amsterdã para a Educação Católica, Cristã Protestante e Interconfessional), conselho que administra 32 escolas na Holanda.

ataque Método para tentar inferir informações sensíveis de indivíduos a partir de uma divulgação de dados.

ataque composicional Ataque que se utiliza do cruzamento de informações divulgadas por mais de uma fonte, ou por divulgações diferentes de uma mesma fonte (e.g., cruzamento de dados de um censo médico com registros públicos de eleitores).

ataque de reconstrução da base de dados Método para reconstruir parcialmente um conjunto de dados privado a partir de informações agregadas públicas.

ataque longitudinal Ataque que se utiliza de informações divulgadas ao longo do tempo (e.g., várias liberações de informação referentes a uma mesma pesquisa, ou pesquisas em anos consecutivos que se referem aos mesmos titulares).

***Australian Bureau of Statistics* (ABS)** Escritório australiano com mandato de coleta de dados estatísticos, incluindo a realização do censo do país.

***Australian Census Longitudinal Dataset* (SCLD)** Base de dados do censo longitudinal australiano, criada a partir de amostras aleatórias com correspondência.

Autoridade Nacional de Proteção de Dados (ANPD) Órgão da administração pública direta federal do Brasil que faz parte da Presidência da República e possui atribuições relacionadas à proteção de dados pessoais e da privacidade e, sobretudo, que deve fiscalizar o cumprimento da LGPD.

banco de dados O mesmo que *base de dados*.

base de dados Conjunto estruturado de dados pessoais, estabelecido em um ou em vários locais, em suporte eletrônico ou físico.

***Bureau of Economic Analysis* (BEA)** Agência federal dos EUA responsável por divulgar dados sobre macroeconomia e indústria, incluindo o produto interno bruto nacional e regional.

***Bureau of Justice Statistics* (BJS)** Agência federal dos EUA responsável por divulgar dados sobre criminalidade.

***Bureau of Labor Statistics* (BLS)** Agência federal dos EUA responsável por divulgar dados sobre economia e trabalho.

***Bureau of Transportation Statistics* (BTS)** Agência federal dos EUA responsável por divulgar dados sobre transporte.

Censo da Educação Básica Principal pesquisa estatística educacional do país, abrangendo a Educação Básica e Profissional.

Censo da Educação Superior Mais completa pesquisa estatística sobre as Instituições de Educação Superior do país.

consulta Uma recuperação de dados ou informações a partir de uma base de dados, normalmente efetuada em linguagem específica.

dado agregado Dado sumarizado correspondendo a uma agregação das informações dos microdados para unidades mais amplas, ou mesmo a uma função computada sobre essas informações (e.g., frequência, média ou outros tipos de análises).

dado anonimizado Dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento.

dado de frequência Tipo de dado agregado usado por agências federais dos EUA relacionado a contagens (e.g., quantos estabelecimentos estão operando em uma determinada região).

dado de magnitude Tipo de dado agregado usado por agências federais dos EUA relacionado a grandezas como lucro (e.g., contagem de estabelecimentos em uma determinada região juntamente com a receita bruta agregada).

dado pessoal Informação relacionada a pessoa natural identificada ou identificável.

dado pessoal sensível Dado pessoal sobre o qual a preocupação quanto à privacidade (e.g., sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico), quando vinculado a uma pessoa natural.

dado tabular Termo técnico específico usado para se referir a um tipo de dado agregado usado por agências federais dos EUA em que sumários sobre microdados são organizados em tabelas.

desanonimização O mesmo que *reidentificação*.

desidentificação Técnica de tratamento de dados pela qual são removidos possíveis identificadores individuais óbvios dos registros (e.g., nome, CPF, RG, ou endereços em níveis mais detalhados que as cidades), na expectativa de dificultar-se uma eventual reidentificação do titular.

differential privacy O mesmo que *privacidade diferencial*.

***Disclosure Review Board* (DRB)** Conselho dos EUA que funciona como um curador que determina a liberação ou não de novos conjuntos de microdados de agências federais do país.

***Economic Research Service* (ERS)** Agência federal dos EUA responsável por divulgar dados sobre agricultura e economia.

***Energy Information Administration* (EIA)** Agência federal dos EUA responsável por divulgar dados sobre energia, incluindo dados sobre carvão, petróleo, gás natural, energia elétrica, renovável e energia nuclear.

Ex. Nac. de Certificação de Competências de Jovens e Adultos (ENCCEJA) Exame realizado pelo Inep que afere competências, habilidades, e saberes de jovens e adultos que não tenham concluído o Ensino Fundamental ou o Ensino Médio na idade adequada.

Exame Nacional de Desempenho dos Estudantes (ENADE) Avaliação do rendimento dos concluintes dos Cursos de Graduação tanto em relação aos conteúdos programáticos previstos nas diretrizes curriculares dos cursos, quanto no que diz respeito às habilidades desenvolvidas e à atualização do aluno em relação ao Brasil e ao mundo.

Exame Nacional do Ensino Médio (ENEM) Avaliação do desempenho escolar ao final da Educação Básica.

***General Data Protection Regulation* (GDPR)** Regulamento do direito europeu de 2018 que trata da privacidade e proteção de dados pessoais, aplicável a todos os indivíduos na União Europeia e do Espaço Econômico Europeu.

informação auxiliar Em relação a uma divulgação de dados de interesse, informação auxiliar é qualquer informação que se possa obter a respeito de um indivíduo através de outros canais que não sejam a divulgação de dados de interesse em si mesma (e.g, via *web*, outros registros públicos, jornais ou revistas, conhecimento prévio).

informação lateral O mesmo que *informação auxiliar*.

Inscrypt *Laboratory of Information Security, Cryptography, Privacy, and Transparency*, laboratório de pesquisa do DCC/UFMG focado em segurança da informação, criptografia, privacidade e transparência.

***Internal Revenue Service, Statistics of Income Division* (IRS)** Agência federal dos EUA responsável por divulgar dados sobre declarações de imposto de renda e informações relacionadas.

k*-*anonymity Método sintático de tratamento que garante que os dados de cada indivíduo pertencente a uma divulgação de dados não possam ser distinguidos dos dados de pelo menos outros $k-1$ indivíduos também pertencentes à mesma divulgação.

l*-*diversity Método sintático de tratamento que aperfeiçoa a técnica de *k-anonymity* ao garantir um nível mínimo de diversidade nos atributos sensíveis de cada conjunto de indivíduos agrupados.

Lei Capiberibe O mesmo que *Lei da Transparência*.

Lei da Transparência Lei brasileira 6.924/2009, que obriga a União, os estados e os municípios a divulgar seus gastos na Internet em tempo real.

Lei de Acesso à Informação (LAI) Lei brasileira 12.527/2011, que trata dos procedimentos a serem observados pelo Estado para garantir o acesso a informações previsto na Constituição Federal do Brasil de 1988.

Lei Geral de Proteção de Dados Pessoais (LGPD ou LGPDP) Lei brasileira 13.709/2018, que dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural.

Marco Civil da Internet (MCI) Lei brasileira 12.965/2014, que regula o uso da Internet no Brasil por meio da previsão de princípios, garantias, direitos e deveres para quem usa a rede, bem como da determinação de diretrizes para a atuação do Estado.

método semântico Qualquer método de tratamento de dados que formule condições semânticas para a divulgação de informação, incorporando em si mesmo a formalização de privacidade almejada (e.g., privacidade diferencial).

método sintático Qualquer método de tratamento de dados que formule condições puramente sintáticas para a divulgação de informação (e.g., desidentificação, *k-anonymity* e *l-diversity*).

microdado A menor fração de um dado e pode estar relacionado a uma pesquisa ou avaliação; o dado em sua menor forma de agregação. Microdados geralmente carregam informação sobre entidades em nível individual (e.g., pessoas, escolas, institutos).

mitigação, técnica ou método de O mesmo que *tratamento de dados*.

***National Agricultural Statistics Service* (NASS)** Agência federal dos EUA responsável por divulgar dados sobre produção agrícola, economia, demografia e meio ambiente.

***National Center for Education Statistics* (NCES)** Agência federal dos EUA responsável por divulgar dados sobre educação e informações sobre finanças de distritos escolares.

***National Center for Health Statistics* (NCHS)** Agência federal dos EUA responsável por divulgar dados sobre saúde pública.

***National Science Foundation* (NSF)** Agência federal dos EUA responsável por divulgar dados sobre educação e pesquisa básica em todos os campos não médicos da ciência e engenharia.

privacidade diferencial Método semântico de tratamento de dados que garante formalmente que apenas uma quantidade negligível de informação sobre um dado indivíduo pode ser obtida a partir dos dados publicados, independemente da participação ou não do indivíduo na pesquisa.

pseudonimização Técnica de tratamento de dados pela qual atribui-se um código individual artificialmente criado a cada registro de uma base de dados, acompanhada da transferência de identificadores diretos (como nome, CPF e RG) para outra base de dados. Dessa forma, o controlador é capaz de mapear identificadores diretos em uma base de dados a registros específicos na outra base através do uso dos códigos individuais, ao mesmo tempo em que se dificultam tentativas de reidentificação do titular por parte de agentes externos.

***Public Use Files* (PUF's)** Tipo de arquivo utilizado pelo USCB (Escritório do Censo dos EUA) que contém todos os registros de todos os entrevistados em uma dada pesquisa.

***Public Use Microdata Samples* (PUMS)** Tipo de arquivo utilizado pelo USCB (Escritório do Censo dos EUA) que contém os registros de apenas uma amostra da população em uma dada pesquisa.

quaseidentificador Informações que não são, em si mesmas, identificadores únicos, mas que estão suficientemente correlacionadas com uma entidade a ponto de poderem ser combinadas com outros quaseidentificadores para criar um identificador exclusivo.

query O mesmo que *consulta*.

registro Informação em uma base de dados correspondente a um indivíduo específico.

Regulamento Geral de Proteção de Dados (RGPD) O mesmo que *General Data Protection Regulation*.

reidentificação Prática de combinar dados anonimizados com informações publicamente disponíveis ou dados auxiliares, a fim de descobrir o indivíduo ao qual os dados pertencem.

***Research Data Centers* (RDC's)** Tipo de sala segura utilizada pelo USCB (Escritório do Censo dos EUA) para acesso controlado e limitado a microdados de pesquisas.

Sistema de Avaliação da Educação Básica (SAEB) Conjunto de testes e questionários aplicados na rede pública e em uma amostra da rede privada que auxilia no diagnóstico da Educação Básica e de possíveis fatores que interfiram no desempenho do estudante.

***Social Security Administration* (SSA)** Agência federal dos EUA responsável por divulgar dados sobre renda e uma lista anual com os nomes mais comumente dados a bebês recém-nascidos no país no ano anterior.

Teorema de Reconstrução da Base de Dados (*Dataset Reconstruction Theorem*) Resultado formal que mostra que, dado acesso a uma quantidade suficientemente grande de informações, alguém pode reconstruir bancos de dados subjacentes e, em teoria, reidentificar indivíduos.

titular Pessoa natural a quem se referem os dados pessoais que são objeto de tratamento.

tratamento (de dados) Toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração.

***United States Census Bureau* (USCB)** Agência federal dos EUA responsável por divulgar dados sobre população e economia.