

Dados do Inep: transparência e proteção da privacidade devem andar juntos

Nota Técnica | 07/2022

O monitoramento do sistema educacional brasileiro se dá principalmente através de avaliações periódicas, pesquisas e a consolidação de índices educacionais realizados pela autarquia ligada ao MEC, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). O Censo Escolar consiste na principal pesquisa estatística educacional, coletando dados do ensino regular e profissionalizante, com participação das escolas públicas e privadas, obtendo informações para a formação dos indicadores educacionais.

Há também o Saeb, avaliação de larga escala com periodicidade a cada dois anos na rede pública e realizado por amostragem na rede particular. Dados dessas duas apurações compõem juntos o principal indicador educacional do Brasil, o Ideb. De grande relevância é também o Exame Nacional do Ensino Médio (Enem), utilizado para avaliar a aprendizagem e as competências dos estudantes ao final do ensino básico.

Como ocorreu em fevereiro, o Inep substituiu os dados históricos do Enem por uma base de dados reduzida, além de divulgar os dados do Enem 2020 e do Censo Escolar de 2021 com certas informações omitidas. Em decorrência disso, parcela dos especialistas em educação apontou que tais bases seriam insuficientes para o

acompanhamento do cenário educacional do país [1]. Como justificativa, o Inep invocou a Lei Geral de Proteção de Dados (LGPD), responsável pela proteção de dados pessoais e privados. A atitude foi embasada em um estudo feito em parceria com a Universidade Federal de Minas Gerais onde foi constatada a possibilidade de identificação individual quebrando assim a anonimidade esperada [2]. Posteriormente, o Inep comunicou a publicação ao longo do ano de 2022 de nova versão das bases de microdados de suas pesquisas anteriores [3][4].

Sobre isso, o Livres:

1. Reafirma que a atuação do Estado deve ser limitada às funções essenciais, protegendo a vida e a liberdade dos cidadãos, e garantindo-lhes o direito a florescer e buscar sua própria felicidade de acordo com suas escolhas e possibilidades.
2. Compreende que o exercício do direito à privacidade dos indivíduos deve ocorrer de maneira adequada e proporcional, garantindo a igual escolha aos demais.
3. Reafirma que a divulgação e os métodos de tratamento dos dados visando a proteção da anonimidade devem ser feitos de maneira transparente e clara.

1. Transparência de dados é pré-requisito para liberdade de escolha educacional

Os dados do Enem, do Censo Escolar, dentre outros censos, estudos e pesquisas do Inep, fornecem informações essenciais para que a sociedade civil possa exercer a necessária fiscalização da qualidade dos serviços públicos educacionais.

Seja fornecida por meio de gestão privada ou estatal, a educação pública é ferramenta de concretização do direito ao florescimento individual, que deve ser garantido de acordo com as escolhas e possibilidades de cada cidadão. Nesse sentido, o acesso transparente a dados sobre a qualidade da prestação dos serviços educacionais é pré-requisito essencial para o exercício da liberdade de escolha educacional e a concretização dos Direitos Humanos.

No plano das políticas públicas, dados socioeconômicos, de raça, distribuição das faixas etárias dos estudantes e outras informações sobre as escolas, estudantes e profissionais da educação são fundamentais para a geração de evidências que viabilizem a construção efetiva de soluções para os desafios de interesse público por meio da pesquisa e análise dos problemas e desigualdades educacionais, em especial no que se refere à avaliação dos impactos da Pandemia de Covid-19 na educação.

Esse monitoramento do quadro educacional e a avaliação das políticas públicas por parte de pesquisadores fornece o embasamento necessário para a tomada de decisões por parte dos gestores públicos e a fiscalização por parte da sociedade civil.

2. É possível aumentar a proteção à privacidade

Antes da decisão do Inep, os dados eram disponibilizados, estritamente falando, de maneira a princípio anonimizada, sem a identificação por nome do estudante ou documento pessoal.

Após a solicitação por várias entidades de documentos que expliquem a decisão do Inep via Lei de Acesso à Informação (LAI), o órgão divulgou nota de esclarecimento[2] onde explica que se embasou em estudos realizados em parceria com a Universidade Federal de Minas Gerais (UFMG) através do Laboratório Inscript (Laboratory of Information Security, Cryptography, Privacy, and Transparency).

No estudo “TED 8750 - PRICE Privacidade nos Censos Educacionais”, detectou-se que o uso de vários identificadores combinados resulta numa possível identificação individual, quebrando assim a anonimidade dos dados e comprometendo sua divulgação de acordo com a LGPD. Segundo o estudo: “[...] somente no Censo Escolar da Educação Básica 2019, o uso de três identificadores (+mês, ano de nascimento e código da escola em que estuda) permite a identificação com probabilidade de acerto de até 29,64%. Se usados quatro identificadores, a chance de sucesso aumenta para 49,86% e, com o uso de todos os dez identificadores, o risco é elevado para 75,51%”. Esse estudo mostra portanto uma preocupação real com a possibilidade de quebra de anonimidade e justifica a necessidade de uma revisão da metodologia do tratamento dos dados divulgados.

Por outro lado, os dados disponibilizados a partir do dia 18 de fevereiro são considerados

insuficientes por parte dos pesquisadores [1][5][6]. Dadas essas considerações, o Livres entende que o caminho intermediário deve ser tomado: os dados devem sim ser remodelados de maneira a corrigir quaisquer vulnerabilidades graves detectadas, mas garantindo a qualidade da informação necessária para a produção de evidências educacionais. Além disso, é importante que o Inep seja claro e transparente com respeito às vulnerabilidades de identificação, às metodologias de tratamento de dados utilizadas e sobre suas tomadas de decisão a respeito.

Como aponta o próprio Inep [2], é possível adotar novas práticas na divulgação dos dados para aumentar a garantia da privacidade do aluno. Neste caso, ajustes podem ser tomados, como a troca do mês de nascimento do estudante por faixas de referência, diminuindo a possibilidade de identificação via cruzamento de dados.

3. Falta de transparência compromete a limitação do Estado pela sociedade

Apesar do esclarecimento prestado e o apontamento dos problemas de identificação pelo Inep em sua nota do dia 22, a divulgação inicial dos dados e a troca das bases antigas foi feita sem comunicação prévia, contrariando a transparência esperada por uma mudança tão importante.

A maneira como a nova forma de publicação dos dados foi disponibilizada inicialmente gera preocupação em outras áreas onde o monitoramento de dados é vital para a formulação de políticas públicas e o seu monitoramento.

Vale também lembrar que a responsabilidade pela educação brasileira deve ser dividida entre os diferentes entes federativos e a sociedade civil, que tem o direito de fiscalizar os resultados das políticas públicas independentemente da gestão de ocasião.

Em uma sociedade livre, o debate público constitui em si mesmo um instrumento de legítima limitação do arbítrio do governante que, para ser efetivo, sempre deve ter a transparência como regra, não exceção.

Conclusão:

O Livres demonstra preocupação com relação à maneira com que a alteração das bases de dados foi inicialmente realizada. Aplicações equivocadas da LGPD podem abrir margem para a falta de transparência dos entes públicos em outras áreas de sua atuação, prejudicando a fiscalização da ação do poder público por parte da sociedade civil.

Ao mesmo tempo, o Livres também considera séria a possibilidade de identificação via o cruzamento da base integral de dados. Tal quebra de anonimidade representaria uma violação da privacidade alheia, expondo assim esses cidadãos ao constrangimento desnecessário.

Nesse sentido, o Livres defende o aprimoramento dos métodos de divulgação dos dados educacionais visando a garantia da privacidade e que as decisões a respeito sejam comunicadas pelo Inep de maneira clara e transparente.

Com a colaboração de

Deborah Bizarria

Coordenadora de Políticas Públicas do Livres

Lucca Radicce

Membro da setorial de Educação do Livres

Mano Ferreira

Diretor de Comunicação do Livres

Referências

1. <https://g1.globo.com/educacao/noticia/2022/02/22/inep-divulga-microdados-do-enem-2020-se-m-informacoes-sobre-escola-e-municipio-dos-participantes-especialistas-criticam.ghtml>. Acesso 22 de fevereiro de 2022.
2. Nota de esclarecimento | Divulgação de microdados. Disponível em: <https://www.gov.br/mec/pt-br/assuntos/nota-de-esclarecimento/divulgacao-dos-microdados>. Publicado em 23/02/2022.
3. <https://www.gov.br/inep/pt-br/assuntos/noticias/institucional/novo-conjunto-de-microdados-e-publicado>. Acesso em 22 de agosto de 2022.
4. <https://www.gov.br/inep/pt-br/assuntos/noticias/institucional/enem-e-enade-tem-novo-conjunto-de-microdados-publicados>. Acesso em 22 de agosto de 2022.
5. <https://jornaldebrasil.com.br/noticias/brasil/inep-exclui-microdados-do-censo-escolar-e-do-enem-e-oculta-informacoes-do-sistema/>. Acesso: 22 de fevereiro de 2022.
6. <https://noticias.r7.com/educacao/inep-exclui-microdados-do-censo-escolar-e-do-enem-anteriores-a-2020-22022022>. Acesso: 22 de fevereiro de 2022.