

INEP

SUGESTÕES PARA APRIMORAR O ACESSO E USO DOS DADOS¹

Sugestões elaboradas pelo Instituto Alfa e Beto e entregues ao Ministro da Educação em 3 de julho de 2019

- A função da avaliação é prover insumos para orientar decisões educacionais, com vistas a promover a melhoria dos resultados.
- O Brasil possui um sistema invejável de testes e avaliações. No entanto, os dados disponíveis não têm se transformado no manancial de informações que seria de se esperar – na forma de relatórios, estudos acadêmicos e orientações para redes de ensino e escolas. A maior evidência disso é o reduzido nível de melhoria da qualidade das notas – ou seja – a avaliação não tem contribuído para melhorar a qualidade (e, muito menos, a eficiência).
- Um dos entraves para que isso ocorra se situa no INEP, tanto na forma de disponibilizar os dados que são públicos quanto na dificuldade de acesso à sala segura do SEDAP.
- O argumento principal utilizado pelo INEP para justificar as dificuldades de acesso aos dados repousa na lei de segurança das informações.
- Em outros países – notadamente nos Estados Unidos, Chile e no âmbito dos projetos levados a cabo pela OCDE, o acesso aos microdados por parte do cidadão e dos pesquisadores é muito mais abrangente, generoso e de menor custo. Isso coloca os pesquisadores brasileiros em enorme desvantagem.
- O objetivo da presente nota é contribuir com sugestões concretas para o INEP, no marco da legislação, no sentido de simplificar, ampliar e promover o acesso cada vez maior, especialmente para os pesquisadores, às suas bases de dados.

¹ Essas sugestões já haviam sido apresentadas ao ex-Presidente do INEP por João Batista Araujo e Oliveira, conforme solicitação do mesmo em reunião realizada no dia 8 de março, na sala da Presidência do órgão.

PROPOSTA 1 – CODIFICAÇÃO DAS VARIÁVEIS IDENTIFICADORAS

Re: Nota do INEP referente ao Processo N. 23036.001038/2019-14

- Por meio da nota supra-mencionada, o INEP informa que irá substituir a identificação de alunos e professores. Com essa medida, o INEP pretenderia ampliar a proteção do sigilo de informações pessoais, dificultando a identificação dos indivíduos constantes nas bases de dados.
 - A implicação dessa medida é que, se implementada, não será mais possível, por meio dos dados disponibilizados via *download*, acompanhar o mesmo aluno e/ou professor ao longo dos anos.
- De acordo com o anunciado pelo INEP, o procedimento será feito de forma retroativa, de tal modo que as bases de dados existentes atualmente serão substituídas.
 - Por conta disso, pesquisadores e instituições que já possuem as bases de dados ficarão em posição privilegiada em relação aos que não as possuem.
- A alegação do INEP se baseia no pressuposto – a nosso ver equivocado – de que a situação atual coloca em risco a proteção do dado pessoal. A nosso ver, isso não ocorre pelas seguintes razões:
 - Não há informações sigilosas no nível do indivíduo nas bases de dados do INEP disponíveis para *download*, tais como nomes, telefones, endereços, número de documentos oficiais (CPF, RG, CNH etc.) ou qualquer outra informação que permita a identificação do indivíduo de forma direta.
 - Somente seria possível identificar um indivíduo com a utilização de outras bases de dados que contenham informações sigilosas e/ou por meio de busca *in loco*. Esses são casos excepcionais que não justificam a restrição de acesso a informação. Ademais, informação sobre os indivíduos também pode ser obtida por outros meios junto às Secretarias de Educação ou escolas – o que escapa totalmente do controle do INEP.
 - Em caso de violação do termo de compromisso e de utilização indevida – com baixíssima probabilidade de ocorrência e mesmo assim em situações muito particulares – não haverá dificuldade em identificar e punir os eventuais infratores, na forma da lei.
 - A experiência do país com o uso controlado da base de dados identificados da RAIS - que inclui nome, CPF, CNPJ etc. – comprova que existem mecanismos adequados para lidar com o sigilo e a responsabilização de eventuais violações – ao amparo da legislação.
- A alternativa proposta pelo INEP – via acesso à sala restrita - não parece ser a melhor solução dadas as dificuldades de acesso à mesma. Mesmo com a adoção de medidas como as sugeridas adiante, isso poderia ser desnecessário se o INEP rever a sua decisão e voltar à situação anterior no que diz respeito ao formato e disponibilização de suas bases de dados.
- Conclusão:
 - Pelas razões expostas, sugerimos que o INEP reveja a sua posição, volte a disponibilizar as bases de dados como vem fazendo ao longo de todos esses anos, e adote a Proposta 2 a seguir.

PROPOSTA 2 – NOVA FORMA DE IDENTIFICAÇÃO DOS ALUNOS

- Situação atual:
 - Atualmente, o INEP fornece um número de identificação para cada aluno no Censo da Educação Básica, que acompanha o aluno ao longo dos anos
 - O mesmo ocorre no Censo da Educação Superior
 - No Saeb e no ENEM, o número de identificação muda a cada ano, mesmo que o aluno seja o mesmo. Dessa forma, não é possível saber o desempenho de um mesmo aluno em provas realizadas em anos diferentes, o que dificulta a realização de inúmeras pesquisas que têm como objetivo, parcial ou total, analisar a evolução do aluno ao longo do tempo, limitando propostas de políticas públicas baseadas em evidência.
- Proposta:
 - O INEP passaria a gerar um identificador único para cada aluno, de tal forma a permitir o cruzamento das bases de dados do próprio INEP
 - O identificador seria o mesmo independente do ano, da instituição em que o aluno está matriculado e da base de dados
 - Por instituição entende-se escola, creche, pré-escola, instituição de ensino superior
 - Por base de dados entende-se Saeb, Censo da Educação Básica, Censo da Educação Superior, ENEM e demais bases produzidas pelo INEP no âmbito nacional, estadual ou municipal
 - O identificador estará presente em todas as bases de dados, para todos os alunos, salvos os casos em que a identificação não seja possível
 - Os números atuais de identificação serão mantidos assim como todas as informações atualmente fornecidas
 - Todas as bases continuariam a ser disponibilizadas para download tal como é hoje (e sem as restrições da Nota do INEP referente ao Processo N. 23036.001038/2019-14)
- Vantagens da proposta:
 - Não há prejuízo no tocante ao sigilo da informação uma vez que i) todas as bases de dados já possuem um identificador para cada aluno e ii) não será adicionada às bases de dados nenhuma informação sigilosa.
 - Ao prover o número de identificação, o INEP aumenta a proteção aos dados sigilosos, uma vez que apenas os técnicos do INEP terão acesso às informações sigilosas que poderão ser utilizadas para a construção do identificador único, tais como nome do aluno, endereço, CPF e RG, entre outros. Atualmente, os usuários interessados em cruzar bases de dados do INEP acessam essas informações na sala segura do SEDAP.
 - A criação do identificador único pelo INEP permitirá que todos os pesquisadores interessados tenham acesso a bases de dados idênticas. Além de uniformizar a informação, o INEP é a instituição com credibilidade para garantir que a informação é confiável.

- A disponibilização das bases de dados com o identificador para *download* permitirá democratizar o acesso, evitando que o acesso a essas informações seja restrito apenas àqueles que possuem recursos financeiros suficientes para acessar as bases de dados na sede do INEP
- O identificador também reduzirá custos no médio prazo, tanto para o INEP quanto para o usuário, uma vez que reduzirá a demanda por uso da sala segura.

PROPOSTA 3 – APRIMORAR ACESSO E USO DA SALA SEGURA DO SEDAP

- Atualmente o INEP propicia o acesso a sala segura. Esse acesso depende de autorização prévia mediante análise da proposta e detalhamento do trabalho e até mesmo dos quadros que serão elaborados.
- O baixíssimo número de pesquisas feitas na sala segura e o nível elementar das pesquisas que puderam ser realizadas com base nas restrições existentes ilustra que o mecanismo conspira contra o uso amplo e adequado das informações existentes para elucidar importantes questões da educação no país.
- O mecanismo atual tem vários inconvenientes:
 - A decisão sobre autorização é dada por pessoas do quadro técnico do INEP, o que, em princípio, possibilita o exercício de critérios administrativos, e não apenas técnico-científicos.
 - O pessoal do INEP tem acesso privilegiado e diferenciado às bases de dados, constituindo, além do privilégio, um potencial conflito de interesses.
 - A demora entre a solicitação, a decisão e os arranjos para viabilizar a consulta ficam limitados a vários critérios, subordinando os pesquisadores às conveniências e condições estabelecidas pelo INEP.
 - Os custos são enormes – exigem deslocamento e permanência em Brasília, e ainda assim sujeita a eventos fortuitos.
 - A sistemática requer que o pesquisador defina a priori os quadros finais que vai obter – o que não constitui boa prática de pesquisa. Normalmente o acesso a base de dados permite um refinamento das hipóteses e dos cruzamentos ao longo do processo – e, para um uso mais adequado dos dados existentes, configura uma forma extremamente inadequada e ineficiente.
 -

Sugestões para aprimorar o acesso a sala segura

Acesso:

- O acesso se dará mediante requerimento do pesquisador especificando os objetivos da pesquisa e as bases que pretende acessar. No caso de pessoa jurídica um Termo de Cooperação pode agilizar ainda mais o acesso dos pesquisadores individuais, pois não haveria necessidade de consulta caso a caso.
- A solicitação será examinada apenas para verificar se os objetivos pretendidos colocam risco de segurança dos dados das pessoas.
- A solicitação será examinada mediante um processo de “blind review” por pesquisadores devidamente qualificados para a tarefa e que não sejam vinculados aos quadros de funcionários do INEP.
- Haverá prazos definidos (poucos dias) para a aprovação e/ou solicitação de esclarecimentos.
- A autorização para o uso da sala se dará em datas acertadas de comum acordo com os pesquisadores-solicitantes.

- Não se faz necessário um documento emitido pela instituição ao qual o usuário é vinculado atestando ciência quanto ao acesso a dados protegidos. Isso torna o processo mais burocrático sem ganho em termos de proteção da informação e exclui do processo pesquisadores autônomos.
- O INEP dará acesso ao “script” gerado pelo usuário durante a manipulação dos dados, permitindo ao usuário justificar os procedimentos adotados e a replicação da pesquisa realizada por parte de outros usuários
- O INEP checará o script antes de encaminhá-lo ao usuário por questões de proteção da informação pessoal
- Afim de não sobrecarregar o INEP, agilizar a pesquisa do usuário e evitar conflito de interesses, e dado que o usuário assina o Termo de Compromisso e Manutenção do Sigilo, não haverá checagem das saídas de resultados pelo INEP e/ou a checagem dar-se-á por meio de uma comissão externa ao INEP.
- Será requisitado do usuário que fizer uso da sala segura o envio de uma cópia eletrônica de todas as publicações que utilizarem os dados acessados, dentro de um período de até um mês após a data da publicação. O não cumprimento do procedimento implicará em sanções ao usuário, que serão aumentadas a cada reincidência.
- As demais regras para a realização da pesquisa permanecem inalteradas desde que não entrem em conflito com o exposto acima

PROPOSTA 4 – AMPLIAR ACESSO A SALAS SEGURAS

O objetivo é ampliar o acesso e reduzir os custos para os pesquisadores.

- **Proposta 4.1.**
 - O INEP conveniar com instituições (ex. IBGE no Rio de Janeiro) que já possuam sala segura para ampliar o acesso dos pesquisadores às mesmas.
 - Nesse caso o pesquisador obedecerá aos requisitos estabelecidos na Proposta 3.
- **Proposta 4.2.**
 - Onde não houver salas seguras de outras instituições: abrir salas seguras, especialmente nos grandes centros.